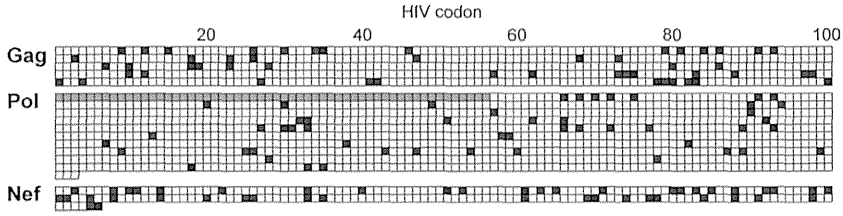

FIG 4 Locations of HLA-associated sites common to HIV-1 clade B-infected Japanese and Caucasian cohorts and those unique to Japan. The locations of all HLA-APs in Gag (500 codons), Pol (1,003 codons), and Nef (206 codons) are illustrated. The residues in the Pol TF protein were not analyzed in the IHAC cohort and are thus excluded (gray bar). The blue squares identify codons that harbored at least one HLA-AP in both Japanese and IHAC cohorts. The red squares indicate codons that harbored HLA-APs in Japan but that were not associated with any HLA alleles in the IHAC cohort.

HLA subtype members varied between HLA allele groups that differed with respect to substitutions within or outside the binding groove, we asked whether the extent of differential escape between subtype members of the former group (comprising A*02, A*26, B*15, B*40, and C*08) differed from those of the latter group (comprising HLA-C*03 and C*14). Overall, we found no significant differences in the proportions of differential escape between them (34.8% for HLA-C*03/C*14 subtypes compared to 36.8% for subtypes of all other HLA alleles; $P = 0.5$) (see Table S2 in the supplemental material). This intriguing result suggests that variations outside the HLA binding groove may contribute as much to differential escape as variations within the binding groove.

Comparison of HLA-APs between Japanese and non-Asian individuals chronically infected with HIV-1 clade B. Our second objective was to investigate HLA-APs identified in Japan versus those previously identified in non-Asian cohorts infected with HIV clade B. The comparison cohort in this analysis was the IHAC cohort, comprising 1,888 antiretroviral-naive individuals with chronic clade B infection in Canada, the United States, and Australia (in which <5% of cohort participants were Asian) (16).

HLA-APs differ to some extent between human populations due to the presence (or enrichment) of certain HLA alleles in one population versus another. Indeed, HLA allele frequencies differed markedly between the Japan and IHAC cohorts (see Fig. S1 in the supplemental material). As such, we began with a qualitative comparison of HLA-APs between them, starting with a simple positional analysis. In the Japanese cohort, HLA-APs were observed at a total of 147 codon positions in Gag, Pol, and Nef (Fig. 4). Of these, 117 (79.6%) were also associated with at least one HLA allele in the IHAC cohort. In contrast, the remaining 30 positions (including 16, 7, and 7 in Gag, Pol, and Nef, respectively) that harbored HLA associations in Japan were not associated with any HLA alleles in the IHAC cohort (Fig. 4). That 30/147 (20.4%) HIV codons exhibited evidence of HLA-driven selection in Japan but not in the IHAC cohort already strongly suggests that HIV is evolving under population-specific selection pressures in Japan compared to other regions.

Next, we compared HLA-APs over HIV position and specific HLA restriction. Of the 284 HLA-APs identified in Japan, 188 (66.2%) were not reported in the IHAC cohort. As expected, a substantial portion of these (46 of 188 [24.5%]) were associated with 8 HLA subtypes (A*26:03, B*40:06, B*54:01, B*55:02, B*59:01, B*67:01, C*08:03, and C*14:03) common in Japan but essentially absent (<1% frequency) in the IHAC cohort. Others were likely attributable to alleles observed at much higher frequencies in Japan than in the IHAC cohort: for example, an additional 27.1% were associated with HLA alleles present in both cohorts but whose frequencies were at least 4-fold higher in Japan than in the IHAC cohort. Overall, the results suggest that HLA-APs identified in Japan are quite distinctive, in large part reflecting the unique HLA allele distribution in the Japanese population.

We also wished to investigate the existence of differential HLA-associated escape pathways between the two populations that are not attributable to HLA frequency differences between them—in other words, cases where the same HLA subtypes drive significantly different escape pathways in the Japan and IHAC cohorts. This required the application of statistical tests (see Materials and Methods and below). Specifically, we first identified a list of 551 HLA-APs in HIV Gag, Pol, and Nef, which represented the union of all HLA-APs identified in either the Japan or IHAC cohort for which both the viral polymorphism and the restricting HLA allele were observed in a minimum of 10 individuals per cohort (not shown). The latter criterion was employed in order to achieve some minimal statistical power to compare the strengths of individual associations between cohorts. It is important to emphasize that these criteria would by definition exclude HLA alleles (and/or viral polymorphisms) present in one cohort but essentially absent in the other (as we would have no power, and in fact no rationale, to test whether their strengths of selection were statistically significantly different between cohorts).

For each HLA-AP, we calculated its lnOR of association in each cohort—a measure that can be interpreted as an estimate of the strength of selection exerted by the HLA allele on that particular HIV codon in that cohort. We then applied a phylogenetically corrected interaction test (17) to assess whether these lnORs of selection were significantly different in the Japanese versus the IHAC cohort. In these analyses, statistical significance was defined as a P value of <0.01 and a q value of <0.05.

Overall, 71 of 551 (12.8%) HLA-APs originally identified in either the Japan or IHAC cohort exhibited significantly different strengths of selection between the two populations (Fig. 5; see Table S3 in the supplemental material). The HLA-B*44:03-associated 125H substitution in Nef serves as an example of how to interpret these data. The lnOR of this association is 1.73 in Japan (with a cohort-specific P value of 3.26×10^{-6}) versus 0.42 for the IHAC cohort (with a cohort-specific P value of 0.36). Both lnORs are positive, indicating that 125H is positively associated with B*44:03 in both cohorts, but the higher lnOR in Japan indicates that the strength of selection of Nef-125H by B*44:03 is greater in Japan than in the IHAC cohort (indeed, the cohort-specific P values reveal that this association is significant in Japan but not in the IHAC cohort). Finally, the P and q values for the intercohort comparison ($P = 1.02 \times 10^{-6}$ and $q = 1.19 \times 10^{-4}$) (see Table S3 in the supplemental material) confirm that the strength of selection

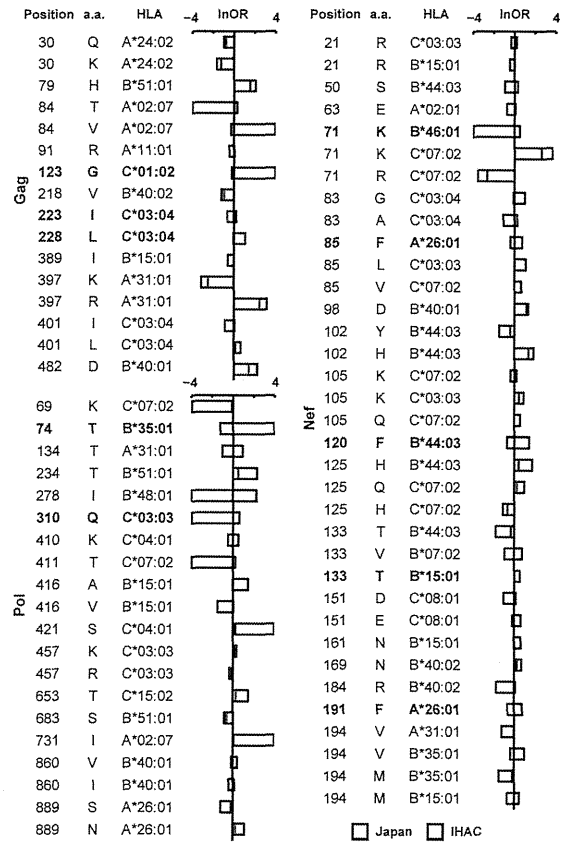

FIG 5 HLA-APs displaying significantly different strengths of selection between Japanese and IHAC cohorts. A phylogenetically corrected interaction test was used to compare the lnOR of selection of HLA-APs in the Japanese cohort versus the IHAC cohort. Comparisons with a P value of <0.01 and a q value of <0.05 are shown. The bars represent the lnORs. Infinite lnORs are set to values of ± 4. Boldface type indicates HLA-APs that display diametrically opposed directions of selection between the cohorts (defined here as lnORs of association that were positive in one cohort but negative in the other, where the cohort-specific P values were <0.05 in both cases). A complete list of all comparisons with a P value of <0.05 is available in Table S3 in the supplemental material.

A

HLA allele	Number of HLA-AP		
	Gag	Pol	Nef
A*02:01	0	0	1
A*02:07	2	1	0
A*11:01	1	0	0
A*24:02	2	0	0
A*26:01	0	2	2
A*31:01	2	1	1
B*07:02	0	0	1
B*15:01	1	2	4
B*35:01	0	1	2
B*40:01	1	2	1
B*40:02	1	0	2
B*44:03	0	0	6
B*46:01	0	0	1
B*48:01	0	1	0
B*51:01	1	2	0
C*01:02	1	0	0
C*03:03	0	3	3
C*03:04	4	0	2
C*04:01	0	2	0
C*07:02	0	2	7
C*08:01	0	0	2
C*15:02	0	1	0
Total	16	20	35

B

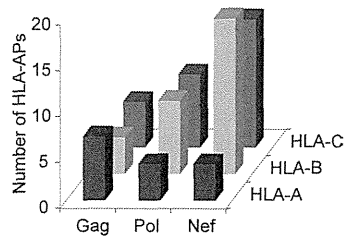

FIG 6 HLA-APs identified as being under differential strengths of selection in Japanese and IHAC cohorts. At a P value of <0.01 and a q value of <0.05, a total of 71 HLA-APs were identified as being under significantly different strengths of selection in the Japanese and IHAC cohorts. (A) Restricting HLA alleles and their HIV-1 protein locations. (B) Numbers of differentially selected HLA-APs, broken down by HLA locus and HIV-1 protein.

of Nef-125H by B*44:03 is significantly greater in Japan than in the IHAC cohort. Importantly, this difference is not simply attributable to intercohort differences in B*44:03 frequencies (which are comparable between populations [see Fig. S1 in the supplemental material]).

In addition to the HLA-B*44:03-associated 125H polymorphism in Nef, we identified 21 other HLA-APs whose strengths of selection were significantly greater in Japan than in the IHAC cohort, yielding a total of 22 (out of 71 [31.0%]) HLA-APs in this category. Conversely, 39 of 71 (54.9%) differentially selected HLA-APs exhibited strengths of selection that were greater in the IHAC cohort than in Japan. The HLA-A*26:01-associated 889S substitution in Pol serves as an example. The lnOR of this association is -0.18 in Japan (with a cohort-specific P value of 0.3) versus -1.17 for the IHAC cohort (with a cohort-specific P value of 7.92×10^{-9}). Both lnORs are negative, indicating that 889S is negatively associated with A*26:01 in both cohorts, but the more negative value for the IHAC cohort indicates that this association is stronger in the IHAC cohort than in Japan. Finally, the P and q values for the intercohort comparison ($P = 1.15 \times 10^{-4}$ and $q = 4.48 \times 10^{-3}$ [see Table S3 in the supplemental material]) confirm that the strength of the negative association between Pol-889S by A*26:01 is significantly greater in the IHAC cohort than in Japan.

Strikingly, the remaining 10 (out of 71 [14.1%]) differentially selected HLA-APs displayed diametrically opposed directions of selection between the cohorts (defined here as lnORs of association that were positive in one cohort but negative in the other, where the cohort-specific P values were <0.05 in both cases) (Fig. 5). The HLA-B*44:03-associated 120F substitution in Nef serves as an example. The lnOR of this association is 1.44 in Japan (with a cohort-specific P value of 2.03×10^{-4}), indicating that HLA-B*44:03 is significantly positively associated with 120F in Japan. In contrast, the lnOR of this association is -0.69 in the IHAC cohort (with a cohort-specific P value of 9.50×10^{-3}), indicating that HLA-B*44:03 is significantly negatively associated with 120F in IHAC. The P and q values for the intercohort comparison ($P = 2.15 \times 10^{-8}$ and $q = 3.75 \times 10^{-6}$ [see Table S3 in the supplemental material]) confirm that the opposing directions of selection of Nef-120F by B*44:03 between the Japanese and IHAC cohorts is a statistically significant observation.

Of interest, the 71 HLA-APs identified as being under significantly different selection in the Japan and IHAC cohorts were differentially distributed across HLA loci and HIV proteins (Fig. 6A and B). Specifically, HLA-A-associated polymorphisms that were significantly differentially selected across cohorts were most abundant in Gag, followed by Pol and Nef, whereas differentially selected HLA-B-associated and HLA-C-associated polymorphisms were most numerous in Nef, followed by Pol and Gag. Taken together, the results support the existence of HLA class I alleles that drive significantly different HIV escape pathways in global populations infected with the same viral clade. The uneven distribution of the locations of these differentially selected polymorphisms across HLA loci and HIV regions raises the intriguing hypothesis that Gag and Pol/Nef may differentially evolve under selection pressures dominated by HLA-A versus HLA-B/C allele-restricted immune responses, respectively.

DISCUSSION

The present study comprised two major objectives, both of which are novel in terms of populations studied and/or analytical methods used. First, we characterized HLA-APs in HIV-1 clade B Gag, Pol, and Nef and their relationship with clinical parameters in a large Japanese cohort. Second, we compared HLA-APs in Japanese versus non-Asian populations infected with HIV clade B to identify population-specific differences in their selection. In particular, we wished to identify HLA-APs that are unique to Japan by virtue of the distinctive HLA distribution in this population, as well as cases where the same HLA allele drives divergent escape pathways in Japan versus non-Asian populations.

This study is the first to identify HLA-APs in HIV-1's structural and functional genes in Japanese populations. Only one previous study investigated HLA-APs in HIV-1 clade B-infected Asians (11): the study comprised 231 Chinese individuals infected during a narrow-source outbreak and identified 141 HLA-associated polymorphisms at two-digit resolution. Our study differs from the previous study with respect to the cohort size, HLA genetics of the host population, HLA-typing resolution, and type of epidemic. Using phylogenetically informed approaches, we identified 284 HLA-APs within HIV-1 Gag, Pol, and Nef in our cohort, supporting a strong influence of population-specific, HLA-driven immune pressures in shaping HIV-1 evolution in Japan. In contrast to a previous study undertaken in a predominantly Caucasian population that observed approximately one-half of the total number of Gag HLA-APs to be located within or flanking reported CTL epitopes (3), the majority of HLA-APs identified in the present study were not located near reported CTL epitopes. This discrepancy may be due to the limited number of Asian-specific HLA-restricted CTL epitopes identified to date, underscoring the need for further epitope discovery in these populations.

This study revealed differential frequencies of HLA-APs across HIV genes in the Japanese population. Consistent with previous studies of HLA-APs in HIV clade B (2, 16, 18), HLA-APs were more frequently detected in Nef than in Gag and Pol. Also consistent with previous observations in Caucasian, African, Chinese, and Mexican populations (1, 6, 11, 15, 18), the number of HLA-B-associated polymorphisms in our cohort was higher than that of HLA-A- or HLA-C-associated polymorphisms, further supporting a dominant role of HLA-B in HIV evolution (32). An interesting feature of the Japanese population is that approximately 70% of individuals carry HLA-A*24:02 (23). Despite sufficient statistical power to detect HLA-A*24:02-associated polymorphisms in our cohort, we identified only 9 of them, 6 of which were located in epitopes identified by our group (33–35). A possible explanation for the relatively low number of A*24:02-associated polymorphisms in Japan is that they have accumulated over time in circulating sequences so that they are no longer significantly enriched among persons expressing HLA-A*24:02. Further analysis of mutations selected by HLA-A*24:02-restricted CTLs should clarify the mechanism whereby high-frequency HLA alleles influence the formation of HIV-1 polymorphisms.

Protective HLA alleles, such as HLA-B*57, -B*58, and -B*27, select Gag mutations affecting viral replication in Caucasians and Africans (36–41) that may also provide some clinical benefit if they are transmitted to hosts lacking these alleles (42, 43). HLA-B*57, -B*58, and -B*27 are not present at appreciable frequencies in Japan (23). It is therefore perhaps unsurprising that no correlations between HLA-associated substitutions in Gag and HIV clinical parameters were observed in our cohort. In contrast, we observed a weak but significant inverse correlation between the frequency of HLA-APs in Pol and the plasma viral load, which appeared to be driven by polymorphisms selected by HLA-B*52:01, an allele identified as protective in Japan (24). Upon further stratification by HLA-B*52:01 expression, the inverse correlation between VL and the total number of B*52:01-associated Pol substitutions was maintained in HLA-B*52:01$^-$ but not in HLA-B*52:01$^+$ individuals. Taken together, these findings suggest that transmitted B*52:01-associated polymorphisms could reduce viral fitness in a dose-dependent manner, though further studies will be required to assess this. In addition, these substitutions were not located within or near known B*52:01-restricted epitopes. Thus, further research will be required to identify these epitopes and elucidate their mechanisms of escape.

Many previous studies of HLA-APs were performed at two-digit HLA resolution (1–4, 6). Here, we performed HLA genotyping at four-digit resolution, which allowed us to investigate differential escape between closely related HLA subtypes in the Japanese cohort. Nearly one-half of the HLA-APs identified in Japan were restricted by HLA allele groups containing two or more subtype members (A*02, A*26, B*15, B*40, C*03, C*08, and C*14). For five of these groups (A*02, A*26, B*15, B*40, and C*08), subtype members differed by substitutions within the peptide-binding groove, while for the remaining two groups (HLA-C*03 and -C*14), subtype members differed by substitutions located outside the peptide-binding groove. Reasoning that amino acid differences located within the peptide-binding groove could modulate the nature or presentation of CTL epitopes, we hypothesized that the former group would generally exhibit distinct HLA-APs between subtype members, while the latter would generally exhibit similar or identical HLA-APs. However, we were surprised to observe substantial evidence for differential HLA-AP selection between closely related HLA subtypes regardless of whether they differed in sequence within or outside the peptide-binding groove. Significantly differential HLA-AP selection was observed at 3 of 9 HLA-C*03-associated sites and 5 of 14 HLA-C*14-associated sites (Fig. 3), proportions that were not significantly lower than the frequency of differential selection between subtypes that differed in their peptide-binding grooves.

This observation raised several hypotheses. HLA polymorphic sites outside the peptide-binding groove may indirectly influence the binding groove conformation, thus altering HLA-peptide interactions and/or T cell recognition. Another possibility is selection by NK cells, as KIR may recognize sites outside the peptide-binding groove. Indeed, KIR3DL1 binds to the loop including position 91 of HLA-B*57:01 (44). However, it is not clear whether KIR2DLs, which are receptors for HLA-C, can bind to the loop outside the peptide-binding groove of HLA-C molecules. A recent study showed that HLA-C antigens are expressed at different levels on the cell surface, even among HLA-C subtypes (45). This study also observed a strong positive correlation between the HLA-C expression level and the strength of HLA-C-mediated selection pressure conferred on HIV. Differential expression levels of these HLA-C subtype members in Japanese populations thus provide another potential explanation for this observation for future follow-up.

Our second objective was to investigate differential HLA-APs between Japanese and non-Asian cohorts infected with HIV clade

B. Here, the IHAC cohort (comprising clade B-infected Canadians, Americans, and Australians) was used as a comparison group (16). HLA-APs identified in human populations differ to some extent due to population-specific HLA distributions, yielding population-specific HLA-APs driven by HLA alleles present in one population but not another (15). Indeed, two-thirds of the HLA-APs identified in Japan had not previously been identified due to the presence of the restricting HLA alleles in Japan but their absence (or far lower prevalence) in the IHAC cohort.

What remains unknown however, is the extent to which the same HLA allele may drive significantly different escape pathways in different human populations. To this end, we applied novel phylogenetically corrected statistical approaches to assess the extent to which HLA-APs identified in either Japan or the IHAC cohort restricted by HLA alleles present in both populations exhibited significantly different strengths of selection. Of the 551 HLA-APs investigated, 71 (12.9%) were significantly differentially selected in Japan versus the IHAC cohort at a stringent statistical threshold of a q value of <0.05. Of these 71, 31% exhibited significantly greater strengths of selection in Japan than in the IHAC cohort, whereas 55% exhibited greater strengths of selection in the IHAC cohort than in Japan. Surprisingly, the remaining 14% displayed diametrically opposed selection pathways in the two cohorts (where an HIV polymorphism represented the adapted form associated with a given allele in one cohort but the nonadapted form associated with the same allele in the other cohort). It is important to emphasize that these significantly different pathways of HLA-AP selection are not simply attributable to differences in HLA frequency between the cohorts.

We feel that these are intriguing observations that merit further study. Nevertheless, we propose the following potential interpretations. First, these differences could be explained by functional differences in HIV-1-specific T cells elicited between the Japanese and Caucasian cohorts, possibly as a result of differences in host genetics (for example, in the genes that encode the T-cell receptor and/or modulate its expression). Such differences may influence the structure of the T-cell receptor(s) and thus the quality, quantity, and/or makeup of the HIV-1-specific T cell repertoire, thus influencing the specific escape mutations selected in the context of peptide-bound HLA. Further analysis of HIV-1-specific T cells driving the selection of these mutants in both cohorts is therefore warranted. It is also important to note that the intercohort HLA-AP comparisons, unlike previous analyses, did not correct for HLA LD or HIV codon covariation. Although both the Japan and IHAC cohorts feature HIV clade B infections, intraclade differences in the viral backbone could also influence differential escape via epistatic effects. In-depth analyses of intercohort differences in HIV codon covariation relationships are therefore also warranted. Intercohort differences in HLA LD are another possible contributor. Finally, the HLA-APs differentially selected between cohorts appeared to be unevenly distributed by HLA locus; while HLA-A-associated polymorphisms exhibiting differential selection between cohorts were more abundant in Gag than in other proteins, HLA-B- and HLA-C-associated polymorphisms exhibiting differential selection between cohorts tended to be more abundant in Nef. This suggests that intercohort differential HLA-APs across HIV proteins may be arising as a result of cellular immune pressures exerted by distinct HLA class I loci, though this also requires further study.

Nevertheless, the present study confirms the existence of population-specific HIV-1 adaptations that are attributable to the unique HLA allele distributions of the population (15). We additionally provide evidence of population-specific HIV adaptation to HLA-restricted immune responses that cannot be explained by differential HLA frequencies alone, cases where the same HLA allele drives significantly different, sometimes opposing, escape pathways in different host populations. Taken together, the results support differential HIV-1 adaptation to human populations worldwide that might be driven by multiple host and viral mechanisms.

ACKNOWLEDGMENTS

We thank Madoka Koyanagi and Rie Maruyama for the collection of samples from patients, Mari Hasegawa and Sayaka Nagata for technical assistance, Sachiko Sakai for her secretarial assistance, and Kyle Cobarrubias for assistance in constructing Fig. 1.

T.C. is a JSPS Research Fellow. A.Q.L. is the recipient of a Frederick Banting and Charles Best Masters award from the Canadian Institutes for Health Research (CIHR). Z.L.B. is the recipient of a New Investigator Award from the CIHR and a Scholar Award from the Michael Smith Foundation for Health Research (MSFHR). This research was supported by the Global COE program Global Education and Research Center Aiming at the Control of AIDS, launched as a project commissioned by the Ministry of Education, Science, Sports, and Culture, Japan, and by Grants-in-Aid for AIDS Research from the Ministry of Health, Labor, and Welfare, Japan.

We have no financial conflicts of interest.

REFERENCES

1. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, Mo T, Hogg RS, Montaner JS, Frahm N, Brander C, Walker BD, Harrigan PR. 2007. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. PLoS Pathog. 3:e94. http://dx.doi.org/10.1371/journal.ppat.0030094.
2. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, Swenson LC, Tao I, Szeto S, Rosato P, Sela J, Kadie CM, Frahm N, Brander C, Haas DW, Riddler SA, Haubrich R, Walker BD, Harrigan PR, Heckerman D, Mallal S. 2009. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. PLoS One 4:e6687. http://dx.doi.org/10.1371/journal.pone.0006687.
3. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, Kadie C, Frahm N, Brander C, Walker B, Heckerman D, Harrigan PR. 2008. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. AIDS 22:1277–1286. http://dx.doi.org/10.1097/QAD.0b013e3283021a8c.
4. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 296:1439–1443. http://dx.doi.org/10.1126/science.1069660.
5. Rolland M, Carlson JM, Manocheewa S, Swain JV, Lanxon-Cookson E, Deng W, Rousseau CM, Raugi DN, Learn GH, Maust BS, Coovadia H, Ndung'u T, Goulder PJ, Walker BD, Brander C, Heckerman DE, Mullins JI. 2010. Amino-acid co-variation in HIV-1 Gag subtype C: HLA-mediated selection pressure and compensatory dynamics. PLoS One 5:e12463. http://dx.doi.org/10.1371/journal.pone.0012463.
6. Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, Prendergast A, Matthews P, Payne R, Rolland M, Raugi DN, Maust BS, Learn GH, Nickle DC, Coovadia H, Ndung'u T, Frahm N, Brander C, Walker BD, Goulder PJ, Bhattacharya T, Heckerman DE, Korber BT, Mullins JI. 2008. HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. J. Virol. 82:6434–6446. http://dx.doi.org/10.1128/JVI.02455-07.
7. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, Schneidewind A, Power KA, Toth I, Frahm N, Alter G, Brander C, Carrington M, Walker BD, Altfeld M, Heckerman D, Allen TM. 2009. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly con-

served regions of human immunodeficiency virus type 1. J. Virol. 83: 1845–1855. http://dx.doi.org/10.1128/JVI.01061-08.

8. Huang KH, Goedhals D, Carlson JM, Brockman MA, Mishra S, Brumme ZL, Hickling S, Tang CS, Miura T, Seebregts C, Heckerman D, Ndung'u T, Walker B, Klenerman P, Steyn D, Goulder P, Phillips R, Bloemfontein-Oxford Collaborative Group, van Vuuren C, Frater J. 2011. Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. PLoS One 6:e19018. http://dx.doi.org/10.1371/journal.pone.0019018.
9. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, Kuse N, Oka S, Duda A, Prendergast A, Crawford H, Leslie A, Brumme Z, Brumme C, Allen T, Brander C, Kaslow R, Tang J, Hunter E, Allen S, Mulenga J, Branch S, Roach T, John M, Mallal S, Ogwu A, Shapiro R, Prado JG, Fidler S, Weber J, Pybus OG, Klenerman P, Ndung'u T, Phillips R, Heckerman D, Harrigan PR, Walker BD, Takiguchi M, Goulder P. 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458:641–645. http://dx.doi.org/10.1038/nature07746.
10. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. 2008. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum. Immunol. 69:443–464. http://dx.doi.org/10.1016/j.humimm.2008.05.001.
11. Dong T, Zhang Y, Xu KY, Yan H, James I, Peng Y, Blais ME, Gaudieri S, Chen X, Lun W, Wu H, Qu WY, Rostron T, Li N, Mao Y, Mallal S, Xu X, McMichael A, John M, Rowland-Jones SL. 2011. Extensive HLA-driven viral diversity following a narrow-source HIV-1 outbreak in rural China. Blood 118:98–106. http://dx.doi.org/10.1182/blood-2010-06-291963.
12. Lau KA, Wang B, Saksena NK. 2007. Emerging trends of HIV epidemiology in Asia. AIDS Rev. 9:218–229.
13. Rodrigo C, Rajapakse S. 2009. Current status of HIV/AIDS in South Asia. J. Glob. Infect. Dis. 1:93–101. http://dx.doi.org/10.4103/0974-777X.56249.
14. Ruxrungtham K, Brown T, Phanuphak P. 2004. HIV/AIDS in Asia. Lancet 364:69–82. http://dx.doi.org/10.1016/S0140-6736(04)16593-8.
15. Avila-Rios S, Ormsby CE, Carlson JM, Valenzuela-Ponce H, Blanco-Heredia J, Garrido-Rodriguez D, Garcia-Morales C, Heckerman D, Brumme ZL, Mallal S, John M, Espinosa E, Reyes-Teran G. 2009. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. Retrovirology 6:72. http://dx.doi.org/10.1186/1742-4690-6-72.
16. Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, Le AQ, Chui CK, Cotton LA, Knapp DJ, Riddler SA, Haubrich R, Nelson G, Pfeifer N, Deziel CE, Heckerman D, Apps R, Carrington M, Mallal S, Harrigan PR, John M, Brumme ZL, International HIVAC. 2012. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. J. Virol. 86:13202–13216. http://dx.doi.org/10.1128/JVI.01998-12.
17. Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, Ndung'u T, Shapiro R, Frater J, Brumme ZL, Goulder PJ, Heckerman D. 2012. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. J. Virol. 86:5230–5243. http://dx.doi.org/10.1128/JVI.06728-11.
18. John M, Heckerman D, James I, Park LP, Carlson JM, Chopra A, Gaudieri S, Nolan D, Haas DW, Riddler SA, Haubrich R, Mallal S. 2010. Adaptive interactions between HLA and HIV-1: highly divergent selection imposed by HLA class I molecules with common supertype motifs. J. Immunol. 184:4368–4377. http://dx.doi.org/10.4049/jimmunol.0903745.
19. Leslie A, Price DA, Mkhize P, Bishop K, Rathod A, Day C, Crawford H, Honeyborne I, Asher TE, Luzzi G, Edwards A, Rousseau CM, Mullins JI, Tudor-Williams G, Novelli V, Brander C, Douek DC, Kiepiela P, Walker BD, Goulder PJ. 2006. Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. J. Immunol. 177:4699–4708.
20. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D. 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput. Biol. 4:e1000225. http://dx.doi.org/10.1371/journal.pcbi.1000225.
21. Carlson J, Kadie C, Mallal S, Heckerman D. 2007. Leveraging hierarchical population structure in discrete association studies. PLoS One 2:e591. http://dx.doi.org/10.1371/journal.pone.0000591.
22. Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100:9440–9445. http://dx.doi.org/10.1073/pnas.1530509100.
23. Itoh Y, Mizuki N, Shimada T, Azuma F, Itakura M, Kashiwase K, Kikkawa E, Kulski JK, Satake M, Inoko H. 2005. High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. Immunogenetics 57:717–729. http://dx.doi.org/10.1007/s00251-005-0048-3.
24. Naruto T, Gatanaga H, Nelson G, Sakai K, Carrington M, Oka S, Takiguchi M. 2012. HLA class I-mediated control of HIV-1 in the Japanese population, in which the protective HLA-B*57 and HLA-B*27 alleles are absent. J. Virol. 86:10870–10872. http://dx.doi.org/10.1128/JVI.00689-12.
25. Sette A, Sidney J. 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50:201–212. http://dx.doi.org/10.1007/s002510050594.
26. Reche PA, Reinherz EL. 2007. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. Methods Mol. Biol. 409: 163–173. http://dx.doi.org/10.1007/978-1-60327-118-9_11.
27. Sidney J, Southwood S, Sette A. 2005. Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. Immunogenetics 57:393–408. http://dx.doi.org/10.1007/s00251-005-0004-2.
28. Sidney J, Peters B, Frahm N, Brander C, Sette A. 2008. HLA class I supertypes: a revised and updated classification. BMC Immunol. 9:1. http://dx.doi.org/10.1186/1471-2172-9-1.
29. Kostyu DD, Hannick LI, Traweek JL, Ghanayem M, Heilpern D, Dawson DV. 1997. HLA class I polymorphism: structure and function and still questions. Hum. Immunol. 57:1–18. http://dx.doi.org/10.1016/S0198-8859(97)00175-4.
30. Boyington JC, Motyka SA, Schuck P, Brooks AG, Sun PD. 2000. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. Nature 405:537–543. http://dx.doi.org/10.1038/35014520.
31. Fan QR, Wiley DC. 1999. Structure of human histocompatibility leukocyte antigen (HLA)-Cw4, a ligand for the KIR2D natural killer cell inhibitory receptor. J. Exp. Med. 190:113–123. http://dx.doi.org/10.1084/jem.190.1.113.
32. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 432:769–775. http://dx.doi.org/10.1038/nature03113.
33. Fujiwara M, Tanuma J, Koizumi H, Kawashima Y, Honda K, Mastuoka-Aizawa S, Dohki S, Oka S, Takiguchi M. 2008. Different abilities of escape mutant-specific cytotoxic T cells to suppress replication of escape mutant and wild-type human immunodeficiency virus type 1 in new hosts. J. Virol. 82:138–147. http://dx.doi.org/10.1128/JVI.01452-07.
34. Koizumi H, Iwatani T, Tanuma J, Fujiwara M, Izumi T, Oka S, Takiguchi M. 2009. Escape mutation selected by Gag28-36-specific cytotoxic T cells in HLA-A*2402-positive HIV-1-infected donors. Microbes Infect. 11:198–204. http://dx.doi.org/10.1016/j.micinf.2008.11.005.
35. Ikeda-Moore Y, Tomiyama H, Miwa K, Oka S, Iwamoto A, Kaneko Y, Takiguchi M. 1997. Identification and characterization of multiple HLA-A24-restricted HIV-1 CTL epitopes: strong epitopes are derived from V regions of HIV-1. J. Immunol. 159:6242–6252.
36. Altfeld M, Kalife ET, Qi Y, Streeck H, Lichterfeld M, Johnston MN, Burgett N, Swartz ME, Yang A, Alter G, Yu XG, Meier A, Rockstroh JK, Allen TM, Jessen H, Rosenberg ES, Carrington M, Walker BD. 2006. HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8(+) T cell response against HIV-1. PLoS Med. 3:e403. http://dx.doi.org/10.1371/journal.pmed.0030403.
37. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, Desouza I, Ryvkin F, Derdeyn CA, Allen S, Hunter E, Mulenga J, Goepfert PA, Walker BD, Allen TM. 2007. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J. Virol. 81:12608–12618. http://dx.doi.org/10.1128/JVI.01369-07.
38. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, Altfeld M, O'Brien SJ, Carrington M. 2005. AIDS

restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. Nat. Med. 11:1290–1292. http://dx.doi.org/10.1038/nm1333.

39. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, Conlon C, Workman C, Shaunak S, Olson K, Goulder P, Brander C, Ogg G, Sullivan JS, Dyer W, Jones I, McMichael AJ, Rowland-Jones S, Phillips RE. 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. 193:375–386. http://dx.doi.org/10.1084/jem.193.3.375.
40. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, St John A, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJ. 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat. Med. 10:282–289. http://dx.doi.org/10.1038/nm992.
41. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, Kuntzen T, Tung CS, LaBute MX, Mueller SM, Harrer T, McMichael AJ, Goulder PJ, Aiken C, Brander C, Kelleher AD, Allen TM. 2007. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J. Virol. 81:12382–12393. http://dx.doi.org/10.1128/JVI.01543-07.
42. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, Treurnicht F, de Rosa DA, Hide W, Karim SA, Gray CM, Williamson C, CAPRISA 002 Study Team, . 2008. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog. 4:e1000033. http://dx.doi.org/10.1371/journal.ppat.1000033.
43. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J, Kaslow RA, Bansal A, Yusim K, Heckerman D, Mulenga J, Allen S, Goulder PJ, Hunter E. 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J. Exp. Med. 205:1009–1017. http://dx.doi.org/10.1084/jem.20072457.
44. Vivian JP, Duncan RC, Berry R, O'Connor GM, Reid HH, Beddoe T, Gras S, Saunders PM, Olshina MA, Widjaja JM, Harpur CM, Lin J, Maloveste SM, Price DA, Lafont BA, McVicar DW, Clements CS, Brooks AG, Rossjohn J. 2011. Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. Nature 479:401–405. http://dx.doi.org/10.1038/nature10517.
45. Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, Yuki Y, Del Prete GQ, Goulder P, Brumme ZL, Brumme CJ, John M, Mallal S, Nelson G, Bosch R, Heckerman D, Stein JL, Soderberg KA, Moody MA, Denny TN, Zeng X, Fang J, Moffett A, Lifson JD, Goedert JJ, Buchbinder S, Kirk GD, Fellay J, McLaren P, Deeks SG, Pereyra F, Walker B, Michael NL, Weintrob A, Wolinsky S, Liao W, Carrington M. 2013. Influence of HLA-C expression level on HIV control. Science 340:87–91. http://dx.doi.org/10.1126/science.1232685.

OPEN ACCESS Freely available online

PLOS | ONE

Phylodynamic Analysis Reveals CRF01_AE Dissemination between Japan and Neighboring Asian Countries and the Role of Intravenous Drug Use in Transmission

Teiichiro Shiino[1], Junko Hattori[2], Yoshiyuki Yokomaku[2], Yasumasa Iwatani[2,3], Wataru Sugiura[2,3]*, Japanese Drug Resistance HIV-1 Surveillance Network[¶]

1 Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo, Japan, **2** Department of Infectious Diseases and Immunology, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan, **3** Department of AIDS Research, Nagoya University Graduate School of Medicine, Nagoya, Japan

Abstract

Background: One major circulating HIV-1 subtype in Southeast Asian countries is CRF01_AE, but little is known about its epidemiology in Japan. We conducted a molecular phylodynamic study of patients newly diagnosed with CRF01_AE from 2003 to 2010.

Methods: Plasma samples from patients registered in Japanese Drug Resistance HIV-1 Surveillance Network were analyzed for protease-reverse transcriptase sequences; all sequences undergo subtyping and phylogenetic analysis using distance-matrix-based, maximum likelihood and Bayesian coalescent Markov Chain Monte Carlo (MCMC) phylogenetic inferences. Transmission clusters were identified using interior branch test and depth-first searches for sub-tree partitions. Times of most recent common ancestor (tMRCAs) of significant clusters were estimated using Bayesian MCMC analysis.

Results: Among 3618 patient registered in our network, 243 were infected with CRF01_AE. The majority of individuals with CRF01_AE were Japanese, predominantly male, and reported heterosexual contact as their risk factor. We found 5 large clusters with ≥5 members and 25 small clusters consisting of pairs of individuals with highly related CRF01_AE strains. The earliest cluster showed a tMRCA of 1996, and consisted of individuals with their known risk as heterosexual contacts. The other four large clusters showed later tMRCAs between 2000 and 2002 with members including intravenous drug users (IVDU) and non-Japanese, but not men who have sex with men (MSM). In contrast, small clusters included a high frequency of individuals reporting MSM risk factors. Phylogenetic analysis also showed that some individuals infected with HIV strains spread in East and South-eastern Asian countries.

Conclusions: Introduction of CRF01_AE viruses into Japan is estimated to have occurred in the 1990s. CFR01_AE spread via heterosexual behavior, then among persons connected with non-Japanese, IVDU, and MSM. Phylogenetic analysis demonstrated that some viral variants are largely restricted to Japan, while others have a broad geographic distribution.

Citation: Shiino T, Hattori J, Yokomaku Y, Iwatani Y, Sugiura W, et al. (2014) Phylodynamic Analysis Reveals CRF01_AE Dissemination between Japan and Neighboring Asian Countries and the Role of Intravenous Drug Use in Transmission. PLoS ONE 9(7): e102633. doi:10.1371/journal.pone.0102633

Editor: Yury E. Khudyakov, Centers for Disease Control and Prevention, United States of America

Received November 23, 2013; **Accepted** June 21, 2014; **Published** July 15, 2014

Funding: This work was supported by a Grant-in-Aid for AIDS research from the Ministry of Health, Labour, and Welfare of Japan [H22-AIDS-004]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: wsugiura@nnh.hosp.go.jp

¶ Membership of the Japanese Drug Resistance HIV-1 Surveillance Network is provided in the Acknowledgments.

Introduction

Since the first HIV-1-infected case was identified in Japan in 1985, the cumulative number of reported cases of HIV/AIDS has been increasing every year, reaching 18,447 by the end of 2010 [1]. The major HIV-1 subtype in the 1980s in Japan was subtype B [2] followed by CRF01_AE [3]. CRF01_AE caused an outbreak among the high-risk heterosexual population in Thailand in the late 1980s [4–6], and was subsequently disseminated to various risk populations in neighboring countries, including Vietnam, Cambodia, Malaysia, Indonesia and China [7–15]. Overall the CRF01_AE is substantial, accounting for an estimated 36% of HIV in South, Southeast, and East Asia (Los Alamos database) in CRF01_AE was likely introduced into Japan's heterosexual population in the early phase of the epidemic [3,4,7], but the characteristics of the spread of CRF01_AE in Japan have not been extensively investigated. Our surveillance research showed that from 2003 to 2008 CRF01_AE was the second most prevalent subtype (8.4%) after subtype B, and its host characteristics are distinct from those of the subtype B population [16]. CRF01_AE cases are significantly linked to heterosexual transmission [3,16,17] and non-Japanese people [16]. In contrast, subtype B tends to be found in men who have sex with men (MSM) and Japanese people.

CRF01_AE cases appear to be diagnosed in Japan at a later stage of infection [16], and trends in the CRF01_AE epidemic in Japan have been difficult to study by conventional descriptive epidemiological methods. However, recent advances in computational science have allowed us to infer the evolutionary dynamics

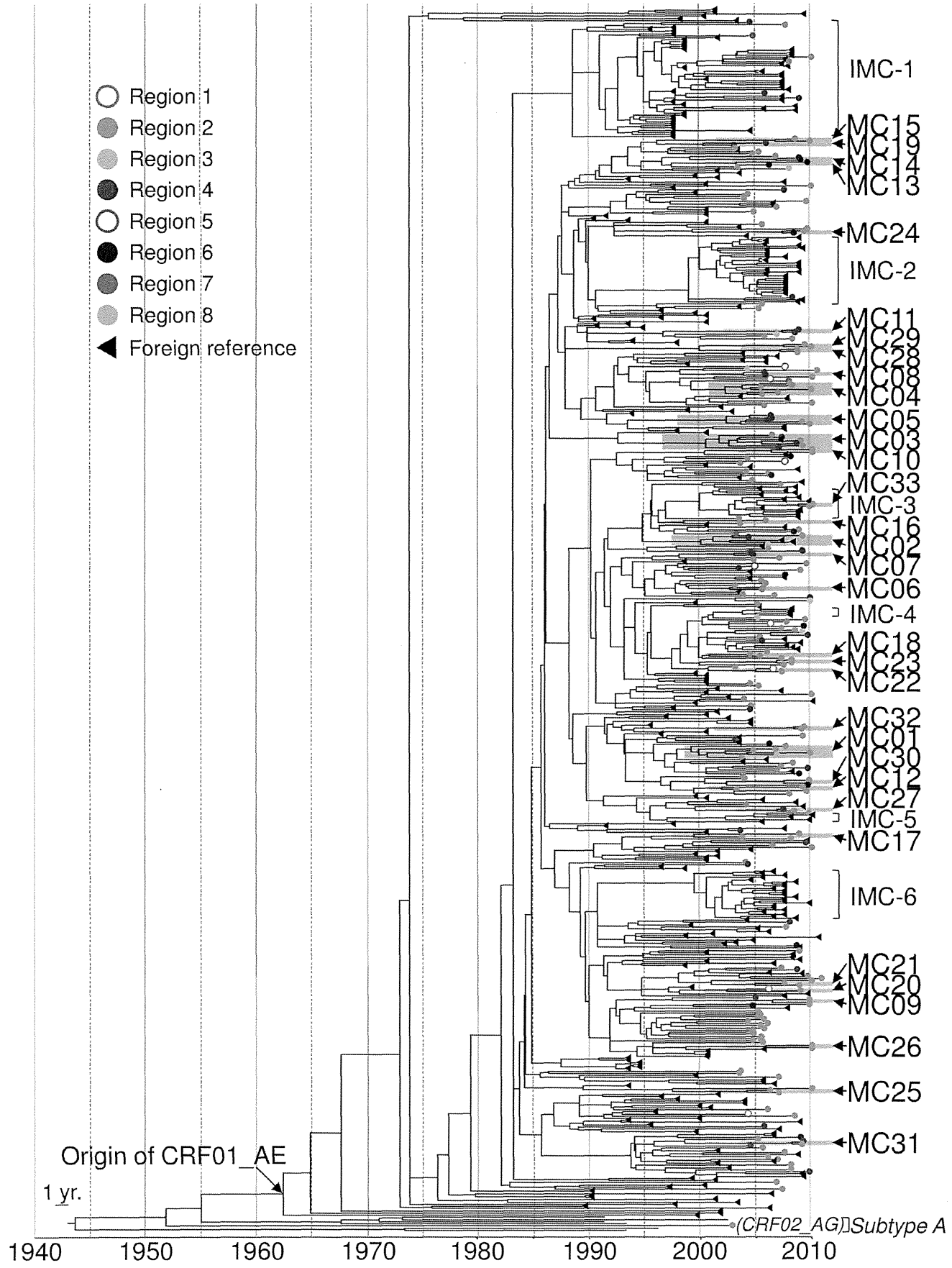
Region 1
Region 2
Region 3
Region 4
Region 5
Region 6
Region 7
Region 8
Foreign reference
IMC-1
MC15
MC19
MC14
MC13
MC24
IMC-2
MC11
MC29
MC28
MC08
MC04
MC05
MC03
MC10
MC33
IMC-3
MC16
MC02
MC07
MC06
IMC-4
MC18
MC23
MC22
MC32
MC01
MC30
MC12
MC27
IMC-5
MC17
IMC-6
MC21
MC20
MC09
MC26
MC25
MC31
Origin of CRF01_AE
1 yr.
(CRF02_AG)Subtype A
1940
1950
1960
1970
1980
1990
2000
2010

Figure 1. Maximum clade credibility tree for partial *pol* region identifies 33 micro-clades within CRF01_AE cases in Japan. Phylodynamic Analysis. The branch length of the phylogeny is in units of time. Patient sequences obtained from surveillance are designated by open or solid circles. Circle color indicates geographic origin of the samples in Japan. Reference sequences from the Los Alamos HIV database are designated by black triangles. Micro-clades (MC) are annotated by red shading and international micro-clades (IMC) are marked with red brackets on the right of the tree.
doi:10.1371/journal.pone.0102633.g001

of a pathogen population from large-scale sequence data using methods, now referred to as "phylodynamics" [18]. Phylodynamics has been used to aid in the analysis of spread of infectious agents with a rapid evolutionary rate [18], e.g., RNA viruses including influenza A [19–21], hepatitis C [22,23], and HIV-1 [24–26]. Since 2003, we have been collecting HIV-1 nucleotide sequence data from newly diagnosed patients in Japan as part of our nationwide surveillance project [16,27]. Here we report our results from applying the phylodynamics approach to these sequence data to understand trends in the CRF01_AE outbreak in Japan, genetic relationships between the circulating strains within Japan and strains observed in the surrounding Asian countries, details of their transmission risk factors, and finally to identify the target populations for effective action plans to prevent further transmission of CRF01_AE.

Materials and Methods

Ethical Statement

This study was conducted according to principles in the Declaration of Helsinki. The study was approved by the human subject research committee at the National Institute of Infectious Diseases and Nagoya Medical Center, Japan. All patients provided written informed consent for the collection of samples and subsequent analyses.

Sample Collection and Viral Gene Sequences

Viral samples were collected from HIV-1-infected patients newly diagnosed from January 2003 to March 2010 at 30 clinics and public health centers in Japan that have participated in our Japanese Drug Resistance HIV-1 Surveillance Network [16,27]. These collection areas are classified into 8 regions according to the nationwide systemic network of hospitals in Japan [28] (Figure S1). The study sample comprised 3618 individuals both acutely and chronically infected. At diagnosis or the earliest hospital visit, patients' peripheral blood was drawn into a vacutainer with EDTA added. At the same time, demographic information was collected on age, gender, nationality, and risk behavior. Plasma samples were analyzed for the nucleotide sequences of HIV-1 protease and the 1- to 240-amino acid region of reverse-transcriptase (RT) using the direct sequencing method of RT-PCR products, and the HIV-1 subtype was determined using phylogenetic analysis as reported [16,27]. This analysis showed that 243 individuals were infected with CRF01_AE at least in the protease-RT regions. GenBank accession numbers of the nucleotide sequences are AB356098–AB556499, AB44228–AB442360, and AB863746–AB871315.

Sequence Alignment of Reconstructed Variants

Direct sequence data may contain loci with multiple peaks, including drug-resistance mutation sites. As these ambiguous loci are generally excluded from analysis by phylogenetic programs, and we wanted to use collected sequences to the maximum in further analyses, we separated these multiple nucleotides into individual nucleotides and reconstructed hypothetical sequence variants possessing each nucleotide as follows. Briefly, protease and RT sequences were concatenated and aligned using Clustal W, version 2.0.10 [29]. Then, a consensus sequence was calculated for the alignment, and ambiguous nucleotides were classified into two groups: overlapping and non-overlapping ambiguities. The former shows one polymorphic nucleotide shared with the consensus allele, while the latter shows a fixation of different alleles from the consensus one [26]. The consensus allele found in the overlapping site was adopted for the reconstructed sequence, and the non-overlapping site was segregated into two haplotypes that carried each nucleotide of the ambiguous site. Consequently, the total number of reconstructed sequences became 297.

These sequences were realigned with foreign CRF01_AE outlier sequences selected as follows. Since transmission clusters have been developed according to scale-free networks in many infectious agents [24,26,30,31], we calculated a frequency distribution of the cluster scale-free network for our observed population in Japan ($n = 297$) by Barabasi's model of scale-free networks [32]. This calculation used the barabasi.game function in igraph library of R software [33]. We estimated that 32.5% of the sequences should be involved in one cluster. Then, using a one-sided error range and rejection coefficient of 0.025 and 1.96 for within and outside clusters, respectively, we calculated the necessary size of the outlier sequence dataset for even allocations within and outside clusters and out of clusters as greater than 332.1. Based on this estimate, we randomly selected 333 CRF01_AE sequences from 37 countries in five regions (Africa, North America, South America, Asia and Europe) submitted to the Los Alamos HIV database before 2010 and used them as the foreign outlier dataset (Table S1). The sequences were further aligned with the following 6 subtype A outgroup sequences: A1.UG.92.92UG037, A1.KE.94.Q23_17, A1.AU.03.PS1044_Day0, A1.RW.92.92RW008, A2.CD.97.97CDKTB48, and A2.CY.94.94CY017_41. The resulting alignment was corrected by hand for gaps. The reference and outlier sequences were collected from the Los Alamos HIV database (http://www.hiv.lanl.gov/content/index.html). The final number of sites in the aligned sequence was 1150 bases.

Phylogenetic Inferences of the Viral Gene Sequence

To eliminate the influence of antiretroviral drug treatments on viral evolution, we made a codon-stripped sequence alignment by removing 43 drug resistance-associated codons defined in our previous studies [16,27]. This alignment was used to estimate a matrix of the number of substitutions between each sequence pair by the composite likelihood method [34], and to infer the neighbor-joining (NJ) tree with the interior branch test. The sequence alignment was also used to infer the maximum likelihood tree using the same substitution model described below with a bootstrap test of 500 replicates. In this process, one of the 244 subjects preliminarily classified into CRF01_AE was re-classified into CRF02_AG (Figure 1 and Figure S2). We excluded this subject in the following analyses. The distance matrix was also used to calculate the mean number of base substitutions per site (i.e., genetic diversity) within and between arbitrary subpopulations, and the coefficient of differentiation between subpopulations. Standard error estimates for genetic diversity were obtained by a bootstrap test with 500 replicates. The analyses were conducted using MEGA version 5.0 [35].

Table 1. Demographic Characteristics of CRF01_AE HIV-1-Infected Individuals in Japan ($N = 243$).

Characteristic		Male		Female		Unknown	Total	
		n	%	*n*	%	*n*	*n*	%
Nationality								
Japanese		128	52.7	36	14.8	1	165	67.9
Asian countries		19	7.8	31	12.8	0	50	20.6
China		1	0.4	3	1.2	0	4	1.6
Philippines		0	0.0	1	0.4	0	1	0.4
Vietnam		3	1.2	0	0.0	0	3	1.2
Malaysia		2	0.8	0	0.0	0	2	0.8
Indonesia		4	1.6	4	1.6	0	8	3.3
Thailand		5	2.1	19	7.8	0	24	9.9
Laos		1	0.4	1	0.4	0	2	0.8
Myanmar		3	1.2	3	1.2	0	6	2.5
South American countries		1	0.4	1	0.4	0	2	0.8
Brazil		0	0.0	1	0.4	0	1	0.4
Peru		1	0.4	0	0.0	0	1	0.4
Unspecified		4	1.6	2	0.8	0	6	2.5
Unknown		13	5.3	4	1.6	3	20	8.2
Transmission category								
High-risk heterosexual contact		104	42.8	62	25.5	0	166	68.3
Male-to-male sexual contact		37	15.2	NA	NA	0	37	15.2
Intravenous drug user		7	2.9	2	0.8	0	9	3.7
Unidentified		17	6.6	10	4.1	4	31	12.8
Area of clinics and facilities								
Region 1	(Hokkaido)	3	1.2	3	1.2	1	7	2.9
Region 2	(Kanto)	112	46.1	48	19.8	3	163	67.1
Region 3	(Koushinetsu)	4	1.6	4	1.6	0	8	3.3
Region 4	(Tokai)	26	10.7	14	5.7	0	40	16.5
Region 5	(Hokuriku)	1	0.4	1	0.4	0	2	0.8
Region 6	(Kinki)	15	6.2	2	0.8	0	17	7.0
Region 7	(Kyushu)	2	0.8	1	0.4	0	3	1.2
Region 8	(Okinawa)	2	0.8	1	0.4	0	3	1.2
Age, years								
20–29		33	13.6	21	8.6	0	54	22.2
30–39		36	14.8	26	10.7	0	62	25.5
40–49		49	20.2	11	4.5	0	60	24.7

Table 1. Cont.

Characteristic	Male		Female		Unknown	Total	
	n	%	*n*	%	*n*	*n*	%
50–59	26	10.7	8	3.3	0	34	14.0
60–69	20	8.2	7	2.9	0	27	11.1
>70	1	0.4	1	0.4	0	2	0.8
Unknown	0	0	0	0	4	4	1.6
Total	165	67.9	74	30.5	4	243	100

NA: not available.

doi:10.1371/journal.pone.0102633.t001

Phylodynamic Analysis

The best-fit model for nucleotide substitution was evaluated by the hierarchical likelihood ratio test using PAUP v4.0 [36] with MrModeltest [37], and the general time-reversible model was adopted with gamma-distributed site heterogeneity and invariant sites (GTR+G+I) with four rate categories. Evolutionary parameters, chronological maximum clade credibility phylogeny, and the times of the most recent common ancestors (tMRCAs) were estimated from the sequence alignment using the Bayesian Coalescent Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.7.4 [38]. The sequence was partitioned into 3-codon positions. To select a model for population growth and the molecular clock, we used Bayesian factor comparison [39] with the marginal likelihood estimated by the stepping-stone sampling method [40,41] using preliminary runs of BEAST with MCMC chains of 100 million iterations. Constant population growth and relaxed clock with an uncorrelated lognormal-distribution was adopted as a best-fit model (Table S2). The best-fit parameters were then used in an additional MCMC analysis consisting of 500 million iterations to estimate the evolutionary parameters. The convergence of parameters was inspected using Tracer v.1.5, with uncertainties depicted as 95% highest probability density (HPD) intervals. The effective sample size of each parameter calculated in this inference was above 200. Tree samples in the MCMC were used to generate a maximum clade credibility tree using TreeAnnotator v.1.5.4 with a burn-in of 40000 states.

Identification of Endemic Transmission Clusters

To identify viral transmission clusters, we performed three analyses; those matching all three approaches were recognized as monophyletic groups. The first approach was to evaluate the reliability of tree topology. We selected transmission cluster candidates from identical tree clusters with three different inference methods: NJ, maximum likelihood and maximum clade credibility tree in Bayesian MCMC. The second approach was a test of monophyly using the interior branch test in NJ tree and a posterior probability in Bayesian MCMC, in which significant clusters were determined as having $\geq$95% confidence probability for a target cluster. The third approach considered genetic divergence of the cluster against the whole sequence diversity. The distributions of all pairwise distances in the given phylogeny were calculated, and a specific sub-tree was then identified as a micro-clade if the median value of genetic distance between each pair of sub-tree members was lower than a threshold, determined as the 10^{th} percentile density (median diversity limit being 0.026) [42]. After a depth-first search of a rooted tree with all three approaches, small groups of viruses with clear evidence of common ancestry (posterior probability $\geq$0.95 in Baysian MCMC inference) were detected; we denoted these groups as "micro-clades", as previously described for small groupings of circulating viral variants viruses [21]. Micro-clades were classified in two groups according to their origins; one is a cluster group having its ancestral virus from Japan (domestic micro-clade), and the other is a cluster group having its ancestral virus from foreign countries (international micro-clade). Micro-clades were then classified as transmission clusters with three or more cases, and cluster candidates as heterosexual, MSM or intravenous drug user (IVDU) pairs. The depth-first search of the clusters was computed by in-house scripts written in Perl 5. The tMRCA was estimated for each domestic micro-clade.

Table 2. CRF01_AE domestic micro-clades (MCs) detected in the study population.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	**Median**	**95% HPD**			**Region**	**Age**	**Gender**	**Behavior**	**Nationality**
Transmission Clusters									
MC01	Nov-'96	Sep-'93	–	Oct-'01	Region 2	35	M	Hetero	Japan
					Region 2	42	F	Unknown	Japan
					Region 2	31	M	Hetero	Japan
					Region 3	44	F	Hetero	Thailand
					Region 2	37	F	Hetero	Japan
					Region 2	40	M	Hetero	Japan
MC02	Mar-'00	Dec-'96	–	Aug-'02	Region 3	30	F	Hetero	Japan
					Region 2	29	M	Hetero	Vietnam
					Region 4	21	M	IVDU	Vietnam
					Reference	-	-	Unknown	Taiwan
					Reference	-	-	Unknown	Vietnam
MC03	Mar-'01	Apr-'98	–	Jun-'03	Region 2	26	M	IVDU	Indonesia
					Region 2	47	M	Hetero	Japan
					Region 4	47	M	Hetero	Japan
					Region 2	32	M	IVDU	Indonesia
					Region 4	23	F	Hetero	Indonesia
					Region 4	26	F	Unknown	Indonesia
					Reference	-	-	Unknown	Korea
MC04	Feb-'02	Sep-'99	–	Jan-'04	Region 2	21	F	IVDU	Japan
					Region 2	66	M	Unknown	Japan
					Region 2	45	F	Hetero	Japan
					Region 2	67	M	Hetero	Japan
					Region 2	60	M	MSM	Japan
					Region 2	63	M	Hetero	Japan
MC05	Jun-'02	Feb-'00	–	Apr-'04	Region 2	28	F	Hetero	Japan
					Region 2	56	M	Hetero	Indonesia
					Region 4	24	F	IVDU	Indonesia
					Region 4	34	M	IVDU	Japan
					Region 4	35	M	IVDU	Indonesia
Heterosexual Male-Female Pairs									
MC06	Feb-'03	Aug-'00	–	Nov-'04	Region 2	68	M	Hetero	Thailand
					Region 2	62	F	Hetero	Japan
MC07	May-'03	Oct-'01	–	Apr-'04	Region 4	50	M	Hetero	Unknown

Table 2. Cont.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	**Median**	**95% HPD**			**Region**	**Age**	**Gender**	**Behavior**	**Nationality**
					Region 4	50	F	Hetero	Japan
MC08	Nov-'03	Jan-'02	–	Dec-'04	Region 4	58	M	Hetero	Japan
					Region 4	51	F	Hetero	Japan
MC09	Jan-'06	Oct-'02	–	Jul-'08	Region 2	40	M	Hetero	Myanmar
					Region 2	33	F	Hetero	Myanmar
MC10	Sep-'06	Aug-'03	–	Oct-'08	Region 2	46	M	Hetero	Japan
					Region 2	29	F	Hetero	Japan
MC11	Apr-'07	Jan-'06	–	Dec-'07	Region 4	66	M	Hetero	Japan
					Region 4	65	F	Hetero	Japan
MSM pairs									
MC12	Nov-'01	Apr-'00	–	Nov-'02	Region 2	29	M	MSM	Japan
					Region 2	36	M	MSM	Japan
MC13	Aug-'03	Jan-'01	–	Jun-'05	Region 6	43	M	MSM	Japan
					Region 6	56	M	MSM	Japan
MC14	Jan-'07	Mar-'05	–	May-'08	Region 6	27	M	MSM	Japan
					Region 6	34	M	MSM	Japan
MC15	Feb-'07	Mar-'05	–	Apr-'08	Region 2	46	M	MSM	Japan
					Region 2	40	M	MSM	Japan
Discordant couples with possible missing cases									
MC16	May-'99	Jan-'99	–	May-'01	Region 2	28	M	Hetero	Japan
					Region 2	46	M	Hetero	Japan
MC17	Sep-'99	May-'96	–	Apr-'02	Region 2	49	M	Hetero	Japan
					Region 2	28	M	MSM	Japan
MC18	Sep-'00	Nov-'96	–	Feb-'03	Region 2	41	M	Hetero	Japan
					Region 2	52	M	Hetero	Japan
MC19	Oct-'00	Nov-'98	–	Jun-'02	Region 4	52	M	Hetero	Japan
					Region 4	65	M	Hetero	Japan
MC20	Dec-'00	May-'98	–	Mar-'03	Region 1	32	M	MSM	Japan
					Region 2	58	M	Hetero	Japan
MC21	Apr-'03	Oct-'99	–	Dec-'05	Region 2	25	F	Hetero	Japan
					Region 2	39	M	MSM	Japan
MC22	Apr-'05	Aug-'03	–	Jun-'06	Region 1	35	F	Hetero	Japan
					Region 2	59	F	Hetero	Japan
MC23	Mar-'07	Sep-'05	–	Jan-'08	Region 2	44	F	Hetero	Japan

Table 2. Cont.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	**Median**	**95% HPD**			**Region**	**Age**	**Gender**	**Behavior**	**Nationality**
					Region 2	34	M	IVDU	non-Japanese
MC24	Sep-'07	Dec-'05	–	May-'08	Region 6	26	M	MSM	Japan
					Region 2	25	M	Hetero	Japan
Insufficient information								?	?
MC25	Aug-'04	Sep-'02	–	Oct-'05	Region 2	20s	M	Unknown	Unknown
					Region 2	28	M	MSM	Japan
MC26	Dec-'05	Nov-'02	–	Mar-'08	Region 2	23	M	Unknown	Unknown
					Region 2	31	F	Hetero	Myanmar
MC27	Mar-'07	Apr-'06	–	Jun-'07	Region 2	45	M	Hetero	Japan
					Region 4	44	M	Unknown	Japan
MC28	Apr-'07	Jun-'05	–	Apr-'08	Region 2	35	F	Unknown	Unknown
					Region 2	29	F	Hetero	Thailand
MC29	May-'07	Feb-'05	–	Dec-'08	Region 2	40s	M	Unknown	Unknown
					Region 2	73	M	Hetero	Japan
MC30	Jun-'07	Jun-'05	–	Oct-'08	Region 2	47	M	Unknown	Unknown
					Region 2	35	M	Unknown	Unknown
MC31	Jul-'08	Apr-'07	–	Jan-'09	Region 2	46	M	Hetero	Japan
					Region 4	39	M	Unknown	Japan
MC32	Oct-'08	Oct-'07	–	Jan-'09	Region 2	24	M	Unknown	Unknown
					Region 2	25	M	MSM	Japan
MC33	Feb-'09	Oct-'07	–	Oct-'09	Region 2	23	M	MSM	Japan
					Region 2	23	M	Unknown	Unknown

MC: micro-clade, HPD: highest posterior density, tMRCA: time of most recent common ancestor.

doi:10.1371/journal.pone.0102633.t002

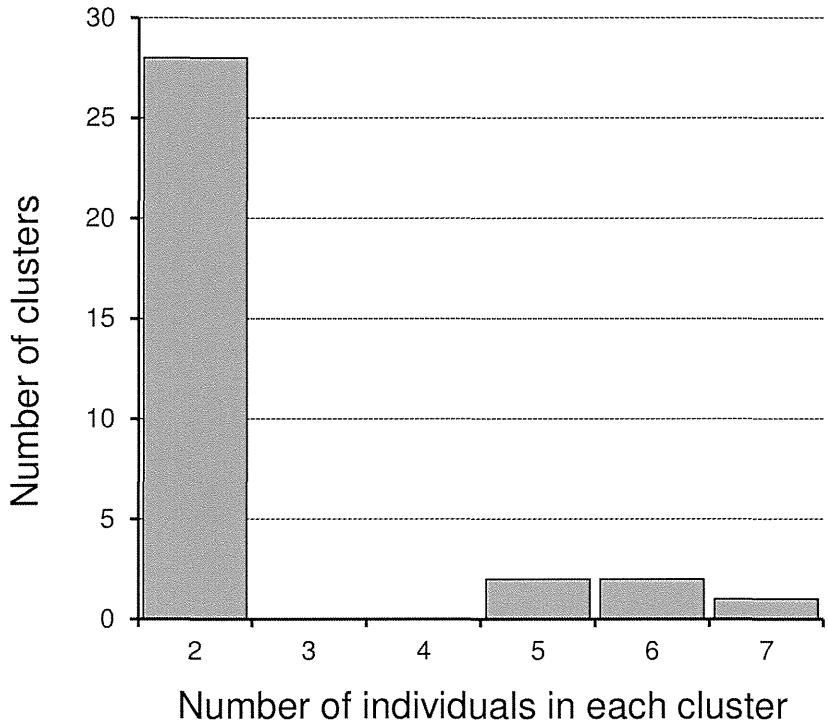

Figure 2. Histogram of CRF01_AE micro-clades by the number of individuals in each micro-clade. Twenty-eight micro-clades consisted of two patients and the remaining 5 micro-clades consisted of more than 4 individuals.
doi:10.1371/journal.pone.0102633.g002

Statistical Analysis

The distribution of IVDUs among micro-clades was tested by χ^2 goodness of fit test to the Poisson distribution. All statistical analyses were calculated using R version 2.10.0.

Results

Major Risk Behavior among CRF01_AE Cases is Heterosexual Contact

We analyzed 297 protease-RT sequences from 243 CRF01_AE-infected cases, which corresponds to 6.7% of the collected study population of 3618 newly diagnosed patients from 2003 to 2010. The highest transmission category in both male and female CRF01_AE populations was high-risk heterosexual contact (Table 1). Among CRF01_AE patients, Japanese males were the majority (52.7%), but this predominance was significantly lower (two sided χ^2 test, $p<0.001$) than in the overall population of HIV infected individuals in Japan (male: 93.2%, Japanese: 90.1%) [16]. The median ages for males, females and the total population were 45, 35 and 40 years, respectively. While patient nationality was not associated with heterosexual behavior (OR = 0.556; $p = 0.183$), MSMs and IVDUs were significantly more (OR = 15.444; $p< 0.001$) and less frequent (OR = 0.086; $p = 0.001$) among Japanese patients, respectively (Table S3).

Phylogenetic Analysis Identifies 33 Micro-clades of CRF01_AE-Infected Patients in Japan

The estimated evolutionary diversity of the CRF01_AE protease-RT region is shown in Table S4. Coefficients of differentiation were low among the risk behaviors and collection areas. The mean evolutionary rate of the protease-RT region estimated by Bayesian MCMC inference was 1.07×10^{-3} substitutions per year (Table S5), consistent with previous estimates for the HIV genome [43–45]. The estimated mean coefficient of variation was 0.597, indicating substantial heterogeneity in the evolutionary rate in viral lineage (Table S5). The tree topology showed no association with any demographic parameters of patients. We identified 33 clusters that mainly spread in Japan, i.e., micro-clades MC01 to MC33 (Figure 1, Figure S2). Some of these micro-clades might have been due to intra-patient sequence variation, as suspected for MC27, 32 and 33. Other micro-clades seemed to come from inter-patient diversity because patients' characteristics were clearly distinct from each other (Table 2). These micro-clades consisted of 76 patients (Table 2), corresponding to 31% of all CRF01_AE patients collected in our study. The distribution of cluster sizes is shown in Figure 2. Most micro-clades (n = 28, 85%) consisted of two patients, with only five micro-clades containing more than a pair.

Origin of the Transmission Clusters Introduced into Japan between 1996 and 2002

The median tMRCAs of micro-clades found in Japan and in foreign reference sequences are shown in Tables 2 and 3, respectively. The median tMRCA of CRF01_AE viruses in Japan was estimated using Bayesian MCMC inference as 1968 (95% HPD: 1975–1956), identical to that of the all CRF01_AE viruses measured in this study, including reference sequences. The earliest transmission cluster originating in Japan was MC01, with tMRCA of 1996. This cluster consisted of 6 individuals (3 males and 3 females), of whom 5 reported their risk behavior as heterosexual contact. Five individuals in MC01 were from Region 2, and one Thai female was detected in Region 3 (Figure 3). Four other transmission clusters (MC02-05) showed median tMRCAs between 2000 and 2002. As shown in Figure 3, these clusters comprised individuals from geographically wide areas of Japan, from eastern to central Japan (Regions 2, 3, 4 and 5; see also

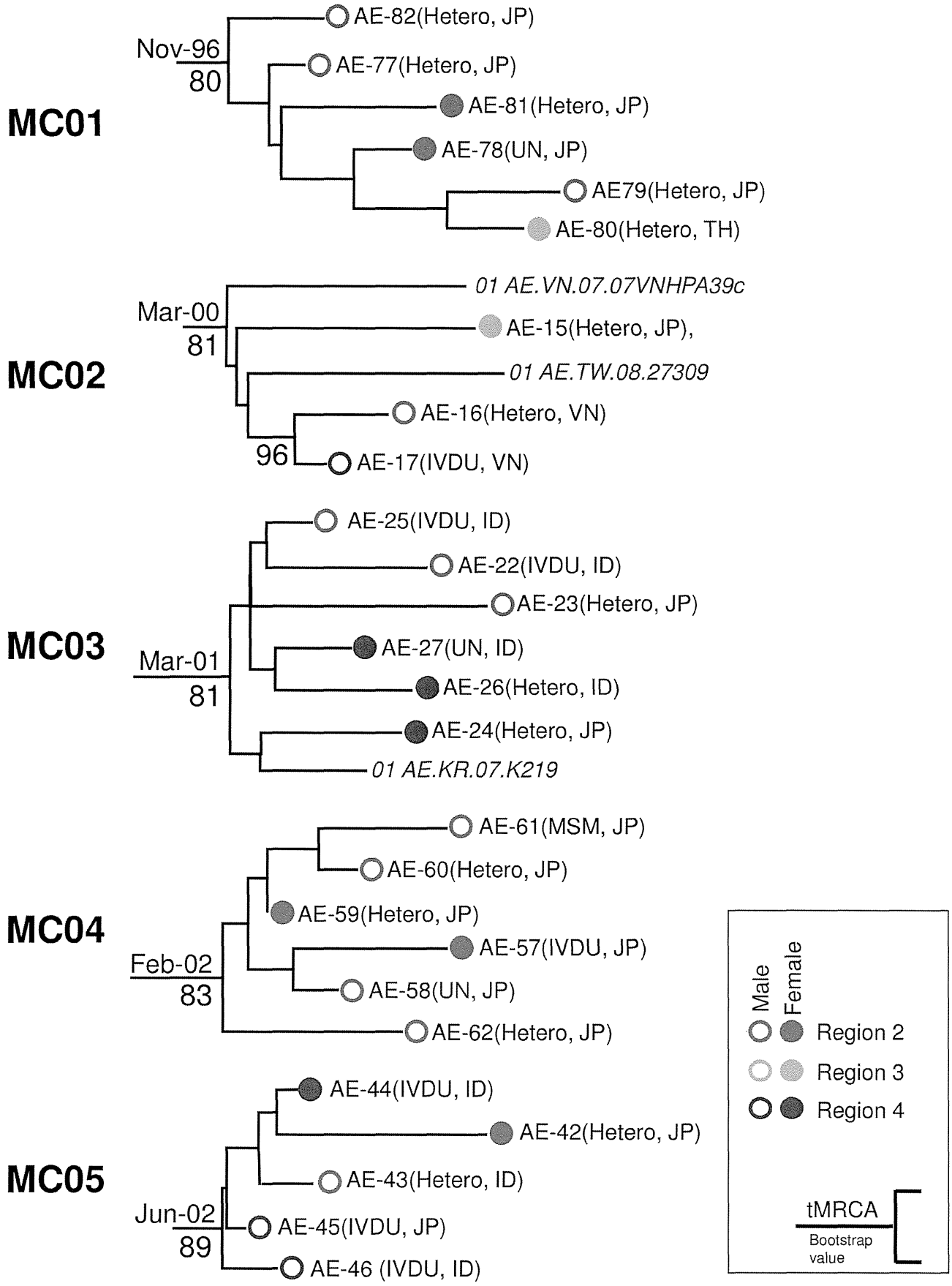
MC01
Nov-96
80
AE-82(Hetero, JP)
AE-77(Hetero, JP)
AE-81(Hetero, JP)
AE-78(UN, JP)
AE79(Hetero, JP)
AE-80(Hetero, TH)
MC02
Mar-00
81
96
01 AE.VN.07.07VNHPA39c
AE-15(Hetero, JP),
01 AE.TW.08.27309
AE-16(Hetero, VN)
AE-17(IVDU, VN)
MC03
Mar-01
81
AE-25(IVDU, ID)
AE-22(IVDU, ID)
AE-23(Hetero, JP)
AE-27(UN, ID)
AE-26(Hetero, ID)
AE-24(Hetero, JP)
01 AE.KR.07.K219
MC04
Feb-02
83
AE-61(MSM, JP)
AE-60(Hetero, JP)
AE-59(Hetero, JP)
AE-57(IVDU, JP)
AE-58(UN, JP)
AE-62(Hetero, JP)
MC05
Jun-02
89
AE-44(IVDU, ID)
AE-42(Hetero, JP)
AE-43(Hetero, ID)
AE-45(IVDU, JP)
AE-46 (IVDU, ID)
Male
Female
Region 2
Region 3
Region 4
tMRCA
Bootstrap value

Figure 3. Detailed phylogenetic structure of five large micro-clades. Open and solid circles indicate male and female cases, respectively. Circle color indicates geographic origin of the samples, followed by risk behavior, and nationality, if known. Dates at the sub-tree's root show the tMRCA of each micro-clade. Numbers below each node show bootstrap values above 80%. JP = Japanese, KR = Korean, VN = Vietnamese, ID = Indonesian, TH = Thai, TW = Taiwanese, UN = unknown, IDVU = intravenous drug user.
doi:10.1371/journal.pone.0102633.g003

Figure S1). These clusters also included both genders and individuals of different nationalities, such as Japanese, Indonesian, Vietnamese, Taiwanese and South Korean, suggesting complex transmission networks in CRF01_AE. The most striking finding was that 7 of 9 IVDUs were significantly concentrated in these four large clusters ($\chi^2 = 528.5$; $p < 2.2 \times 10^{-16}$).

Small CRF01_AE clusters frequently included MSM

In contrast to the large CRF01_AE clusters, which contained only 1 MSM/29 total individuals in clusters, the small CRF01_AE clusters frequently included individuals reporting MSM risk factors. The median tMRCA of small micro-clades ranged between 1999 and 2009 (Table 2). Among these micro-clades, 23 consisted of samples collected in the same region, whereas individuals in the remaining 5 micro-clades were widely distributed over mainland Japan (MC20, 22, 24, 27, and 31). Five micro-clades (MC06, 09, 23, 26, and 28) included at least one non-Japanese. The proportion of females in these small micro-clades was 23%, lower than that of the entire study population (30.5%, Table 1). Six micro-clades (MC06-11) consisted of heterosexual pairs that were unlikely to expand their transmission networks. The median tMRCAs of such closed micro-clades ranged from 2003 to 2007, and those of 4 micro-clades (MC12-15) consisting of MSM pairs ranged from 2001 to 2007. Other 26 small micro-clades had diverse demographic, and risk factor characteristics (Table 2). Taken together with the findings of the large clusters, these data highlight the complexity of CRF01_AE transmission in Japan.

CRF01_AE Epidemic in Japan Occurred in at least Two Waves

The distributions of numbers of individuals in micro-clades and their demographic parameters are graphed versus micro-clade tMRCA in Figure 4. Analysis of tMRCA revealed two distinct groups of CRF01_AE infected individuals: the first wave coming in the early 2000s and the second wave in 2007 to 2008 (Figure 4A). Before the first wave, the major risk factor in the clusters was heterosexual behavior, with a remarkable number of IVDUs around the first wave (Figure 4B). After the first wave, MSM increases gradually and reaches a peak at the second wave (Figure 4B). The clusters in the first wave included many Indonesians (Figure 4C). Thailand, where the CRF01_AE outbreak started, had few contributions to cluster formation in any years, and Japanese mainly contributed in the recent 4 years from 2006 to 2010 (Figure 4C). Individuals from Region 6 were found in the clusters only after 2003, showing that the infection seems to have spread from eastern Japan to the rest of the mainland around this year (Figure 4D). These patterns were not found in analyses of the relationship between demographics and collection year (Figure S3). These data suggest CRF01_AE was introduced at least twice, with one early wave occurring in the early 2000's largely transmitted through heterosexual contact, and a second distinct wave of transmission occurring later that has been sustained by transmission through multiple risk factors, that includes a substantial contribution of MSM transmission.

Some Individuals Infected with Viruses from the Transmission Clusters Spread in East and Southeast Asian Countries

Besides domestic micro-clades, 6 international micro-clades (IMC-1-6) were determined in reference CRF01_AE sequences from the Los Alamos HIV database (Figure 1, Table S1). Among these international micro-clades, 4 (IMCs 1-3 and 5) included sequences from our study (Table 3, Figures S4 and S5). Seven individuals from a wide range of regions in Japan were grouped into IMC-1, the largest international micro-clade with reference sequences collected in Vietnam and China. IMC-2, the second largest cluster consisting of 27 sequences from China, contained one Chinese female detected in Region 4. IMC-3 consisted of 12 sequences from China and included MC33, which included 1 or 2 Japanese males from Region 2. IMC-5, which contained reference sequences from the Philippines, included one MSM Japanese male from Region 2. Thus, our data clearly demonstrate ongoing international transmission of CRF01_AE, especially between China and Japan.

Discussion

The results of this study reveal a process of multiple transmissions and a subsequent spreading pattern of CRF01_AE infection into Japan. Our sample represents 18.7% of the officially reported HIV-1-infected cases in Japan [16]. Although our results were obtained from the relatively short and genetically conserved *pol* sequence region, we collected the largest number of sequences for this region and we considered that this greater number would be more advantageous for our study. The age distribution of CRF01_AE is shifted more toward older populations compared to subtype B cases. However, as non-B subtype HIV-1-infected individuals tend to visit a clinic when they have not recently seroconverted (>155 days) [16], the estimated age at infection may be similar to that for subtype B HIV-infected cases. Unlike subtype B-infected cases, among which MSM is the major risk factor, more than half of the CRF01_AE-infections in Japan were estimated to be transmitted through heterosexual contact.

Using composite phylogenetic analysis, we determined at least 30 clusters of CRF01_AE-infected individuals in Japan; we noted 6 groups had international members, as they contained both our surveillance data and reference sequences from other Asian countries. Among 243 patient samples collected in the study, 76 were from patients involved in these micro-clades, demonstrating transmission networks of the CRF01_AE infection in Japan. The remaining 167 samples failed to be identified in transmission clusters; the CRF01_AE epidemic is quite large, and it is possible that sequences without identifiable partners is due to the low coverage rate of our surveillance or that they were newly imported from neighboring countries.

An original outbreak of CRF01_AE among the high-risk heterosexual population was first reported in Thailand in the late 1980s [4–6], then disseminated to neighboring countries [8–12]. Soon after the outbreak in Thailand, the first extensive colonization of the virus was estimated in Japan. Indeed, our phylodynamic analysis suggests that the primary domestic transmission cluster of CRF01_AE was formed in the 1990s. This possibility was confirmed from IMC-1, which includes 7 domestic sequences

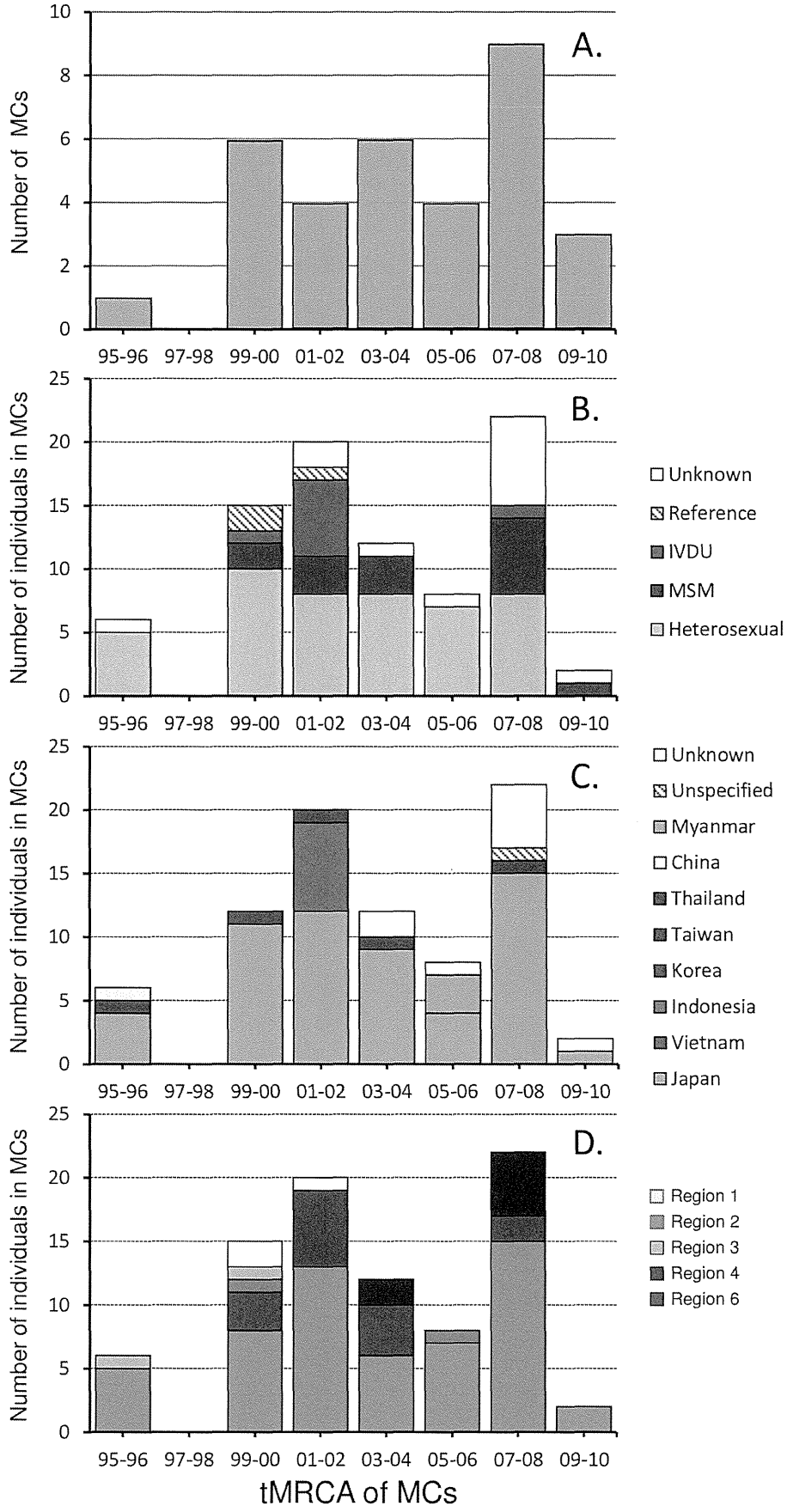

Figure 4. Dynamics of transmission clusters in Japan according to member individuals' demographic characteristics. A) Number of individuals in micro-clades is shown versus micro-clade tMRCA (time of the most recent common ancestor). Bars in each panel are colored by B) risk behavior, C) nationality, and D) geographical region. MC = micro-clade.
doi:10.1371/journal.pone.0102633.g004

Table 3. International micro-clades (IMCs) detected in the study population and outlier sequences.

	tMRCA				Characteristics of individuals in micro-clades				
IMC-ID	**Median**	**95% HPD**			**Region**	**Age**	**Gender**	**Behavior**	**Nationality**
IMC-1	Dec-'94	Mar-'93	–	Mar-'96	Region 7	26	F	Hetero	Japan
					Region 4	24	M	Hetero	Japan
					Region 4	62	M	Hetero	Japan
					Region 2	35	M	MSM	Japan
					Region 2	54	M	Hetero	Japan
					Region 4	43	M	Unknown	Japan
					Region 2	66	M	Hetero	Japan
					Vietnam	35 ref seq (see Table S1)			
					China	7 ref seq (see Table S1)			
					France	2 ref seq (see Table S1)			
					Czech	2 ref seq (see Table S1)			
IMC-2	Oct-'99	Dec-'96	–	Dec-'01	Region 4	32	F	Unknown	China
					China	27 ref seq (see Table S1)			
IMC-3	Feb-'00	Aug-'99	–	Dec-'01	MC33	2 individuals in Japan (see Table 2)			
					China	12 ref seq (see Table S1)			
IMC-4	ND	ND		ND	China	4 ref seq (see Table S1)			
IMC-5	Aug-'02	Jun-'99	–	Feb-'05	Region 2	30	M	MSM	Japan
					Philippines	3 ref seq (see Table S1)			
IMC-6	ND	ND		ND	China	23 ref seq (see Table S1)			
Sequence groups in CRF01_AE									
Japan AE	Nov-'68	Sep-'56	–	Apr-'75	-	-			
Whole AE	Nov-'68	Sep-'56	–	Apr-'75	-	-			

HPD: highest posterior density, tMRCA: time of most recent common ancestor, ND: not determined.

doi:10.1371/journal.pone.0102633.t003

and "cluster 3" containing sequences from North Vietnam and a neighboring Chinese province described by Liao et al. [46], suggesting simultaneous disseminations to these countries, including Japan (Figure S5). Our tMRCA analysis demonstrated that some transmission clusters independently spread in East Asia from around 2000 (Table 3). Among these, IMC-1 was the same as the cluster found in IVDU populations in Northern Vietnam and Southeastern China [46,47]. The estimated median tMRCA of IMC-1 was similar to that of previous studies [46,47], and our subjects from regions 2 and 4 were scattered within the cluster, suggesting repeated transmission events of CRF01_AE from these regions to Japan around 2000. In turn, IMC-3 was the same as a cluster recently found in a northeastern Chinese MSM population (named "cluster 1" [48–50] or CN.MSM-01-01 [51]) and included the domestic micro-clade, MC33.e constructed a maximum likelihood tree using our subjects with 4 CN.MSM-01-01 sequences derived from Japan with the same substitution model and confirmed that members of MC33 were involved in a Japanese sub-cluster of CN.MSM-01-01 (Figure S6). Therefore, one can observe that MC33-related viruses may have contributed an outbreak in an MSM community in metropolitan areas of Japan. Additionally, domestic clusters MC02 and MC03 included reference sequences originating from Taiwan and Vietnam, and South Korea, respectively (Figure 3). Thus our data clearly suggest that transmission networks of CRF01_AE have developed between Japan and other Asian countries from the first colonization wave. The CRF01_AE epidemic scenario would be that CRF01_AE was initially imported to Japan in the 1990s from neighboring Asian countries especially by IVDU behavior. Then, Japanese variants may have influenced epidemics among MSM in East Asian countries, as suggested by Kondo et al [51], by exporting cases in the middle of the 2000s. Among such "connected" countries in Asia, generation of recombinant forms between CRF01_AE and other subtypes has been a recent concern [7,51].

After the primary dissemination to Japan, our tMRCA distribution data strongly suggest that CRF01_AE invaded Japan in two substantial waves, one in 2000 and the other in 2007. Our finding is based on the ability to detect common ancestors among circulating viral variants. We could not assign common ancestors to all of our sequences, and the CRF01_AE epidemic is substantial and genetically diverse. As a result, our sampling, as well as sampling in other countries, may not be sufficient to detect additional, ongoing introductions of CRF01_AE. A serious concern drawn from our findings is the role of IVDU in the CRF01_AE transmission network. The concentration of IVDUs observed in large micro-clades (e.g., MC05) indicates a suspected linkage of high-risk sexual communities and drug addicts, including international partners (MC05) As stigma often limits standard risk assessment for HIV, the phylodynamic approach described here offers an effective method to track ongoing epidemics within suspected IVDU communities, with the results of our analysis indeed clarifying critical contributions of IVDU to the CRF01_AE outbreak in Japan. These results aid in calling attention to the need to focus resources on interventions designed to specifically limit spread among specific risk groups to curtail CRF01_AE transmissions through the IVDU route.

Supporting Information

Figure S1 Geographic location of HIV-1 sample collection regions in Japan. Regions of sample collection are designated by the same colors used to indicate sample origin in other figures.
(PDF)

Figure S2 Distance-based neighbor joining phylogeny of the protease-RT region of CRF01_AE HIV-1 in Japan. Numbers on each branch show the results of interior branch testing, where probabilities >95%. The sequences obtained in our surveillance network are designated by circles in different colors according to the region of sample collection. Reference sequences from the Los Alamos database are designated by black triangles. Micro-clades and significant clusters are annotated by red branches with brackets on the right of the tree. Scale bar at the bottom shows the number of nucleotide substitutions per site.
(PDF)

Figure S3 Distribution of sample collection time of CRF01_AE HIV-1-infected individuals in Japan. The cumulative numbers of CRF01_AE HIV-1-infected individuals are shown by year of sample collection. Bars in each panel are colored by individuals' A) gender and risk behavior, B) nationality, and C) geographical region.
(PDF)

Figure S4 Partial chronological phylogenetic tree of IMC-1. An international micro-clade including 7 sequences from our study population and a cluster that spread mainly in Vietnam [46] extracted from the Bayesian MCMC phylogeny is shown. The 7 sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Figure S5 Partial chronological phylogenetic tree of IMC-3. An international micro-clade composed of CRF01_AE sequences found in China extracted from the Bayesian MCMC phylogeny is shown. This cluster included MC15. Sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Figure S6 Maximum likelihood phylogenetic tree of the large Chinese cluster CN.MSM.01-01 with MC33. Protease-RT sequences belonging to CN.MSM.01-01 [51] were selected from the Los Alamos database and aligned with our study subjects and outlier sequences. Maximum likelihood phylogeny was inferred from the alignment as described in Materials and Methods. A partial tree including CN.MSM.01-01 is represented. Numbers below branches indicate bootstrap probability. Japanese sequences in CN.MSM.01-01 are underlined. Our sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Table S1 CRF01_AE outlier sequences from the Los Alamos HIV database.
(PDF)

Table S2 Bayesian factor analysis of molecular clock models compared for constant demographic size.
(PDF)

Table S3 Independence test of risk behaviors and nationalities in CRF01_AE-infected patients in Japan.
(PDF)

Table S4 Estimates of the mean evolutionary diversity for categories of CRF01_AE sequences.
(PDF)