tions of the A3F-Vif interaction, because no information is available concerning the structure of the A3F-Vif (or A3F-Vif-CBF-β) complex to date.

The A3F-binding interface is clearly separate from the YRHHY (residues 40 to 44) position required for A3G binding (Fig. 2). Detailed analysis of the A3F-binding interface conformation using a Vif structural model with a minimum extension of RWNKP (residues 173 to 177) exhibited two unique characteristics: (i) one-half of the region responsible for A3F binding, especially a portion consisting of R17, E171, and R173, represents a highly positively charged surface, and (ii) a hydrophobic side chain of M16 projects out of the surface (Fig. 2). Interestingly, the electrostatic distribution of the Vif interface is inversely correlated with the electrostatic distribution of the negatively charged A3F interface. Additionally, our analysis of Vif mutagenesis demonstrates that the Vif interface is more extended for A3F than for A3C (Fig. 1 and 2), which coincides with the evidence that four additional residues of A3F are critical for the Vif interaction. These data suggest that the Vif surface composed of R17, E171, and R173 might electrostatically interact with the extended interface region, including A3F residues A292, R293, and E324. In a parallel manner, the hydrophobic M16 of Vif might directly interact with the hydrophobic A3F cavity between the α2 and α3 helices because these hydrophobic interactions are critical for the degradation of A3F as well as A3C. Therefore, we sought to predict a model of the A3F CTD and Vif complex using docking simulations, with Vif M16 fixed in the hydrophobic A3F cavity (Fig. 8; see Movie S1 in the supplemental material). The model structure suggests that Vif R17, E171, and R173 and A3F A292, R293, and E324 tend to cluster around the bottom region between the A3F α3 and α4 helices, whereas Vif W174 is likely located at the A3F CTD cavity between the α2 and α3 helices. These observations support our proposal that the electrostatic complementarities between the A3F-unique residues and concurrent anchoring of Vif hydrophobic residues into A3F are required for the A3F-Vif interaction. Moreover, our detailed analyses of the structural model suggest (i) the model is plausible, (ii) two Vif residues (M16 and W174) are clustered to form hydrophobic interaction with the shallow cavity (L255, F258, C259, I262, Y269, F290 and H294) of A3F CTD α2/α3, (iii) three hydrophilic residues (E289, R293 and E324) clustered at the A3F CTD α3/α4 region are positioned closely to four Vif residues (D14, R15, E76, and R173), and (iv) four Vif residues (R17, E171, D172, and N175) of the F-box motifs are proximal to each other and are likely to form intramolecular hydrogen bonds and/or salt bridges. These residues (except for N175) involved in their interaction in our structural model were actually identified by our experiments (Fig. 1, 3, and 4) as critical residues for the A3F-Vif interaction, supporting our structural complex model. In contrast, we identified a new potential interaction that was not investigated: Vif W79 (F2 box) is close to A3F Y314. Therefore, we further assessed whether the Y314 plays a critical role in the Vif-mediated degradation. Three additional A3F mutants, A3F Y314A, Y314E, and Y314K, were constructed, and their Vif sensitivity was evaluated. The results demonstrated that the three A3F mutants (Y314A, Y314E, and Y314K) exhibited Vif sensitivity, suggesting there was no involvement of the A3F Y314 residue in their interaction. The mechanistic reason for this discordance regarding Y314 is not clear yet. Further experiments are required for refinement of our structural model.

The Vif F-box interface consists of the flexible loop structures

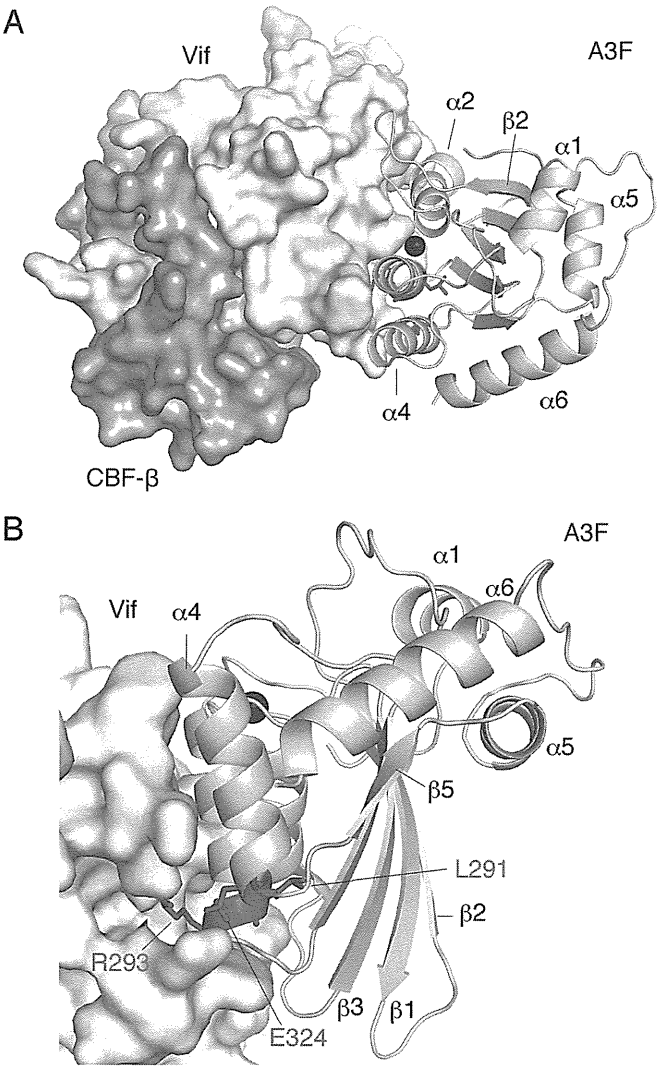


FIG 8 Predicted interaction between the A3F CTD and Vif. (A and B) Two views of their interaction surfaces, from the A3F catalytic groove side (A) and from the A3F CTD α4 helix side (B), are shown. Structures of the A3F CTD and Vif are represented as wheat ribbons and gray surface, respectively. The A3F-unique residues crucial for Vif binding are highlighted with magenta sticks. A zinc ion in the catalytic groove is shown as a black sphere.

of the F1, F2, and F3 boxes. This plasticity might be of adaptive advantage for viruses to change substrate recognition interfaces with overall Vif architecture maintained, especially when viruses need to antagonize different species of A3F. Nevertheless, the structural features of the A3F-binding interface on Vif are likely maintained among the major group of HIV-1 (HIV-1 M) and chimpanzee SIV (SIVcpz), as the interface residues critical for the A3F interaction are highly conserved among Vif sequences (Fig. 9). It has been reported that during adaptation of SIVs in chimpanzees, a unique 3′-*vif* region containing the entire F3 box was created by overprinting associated with the *vpx* gene loss (63). Such a virus adaptation might be due to high selective pressures that forced HIV-1/SIVcpz Vif to extend the A3F-binding interface to maintain the Vif-mediated antagonizing activity against the A3F protein with high antiviral activity compared with A3C (64, 65). Chimpanzee A3D might also play an important role in the

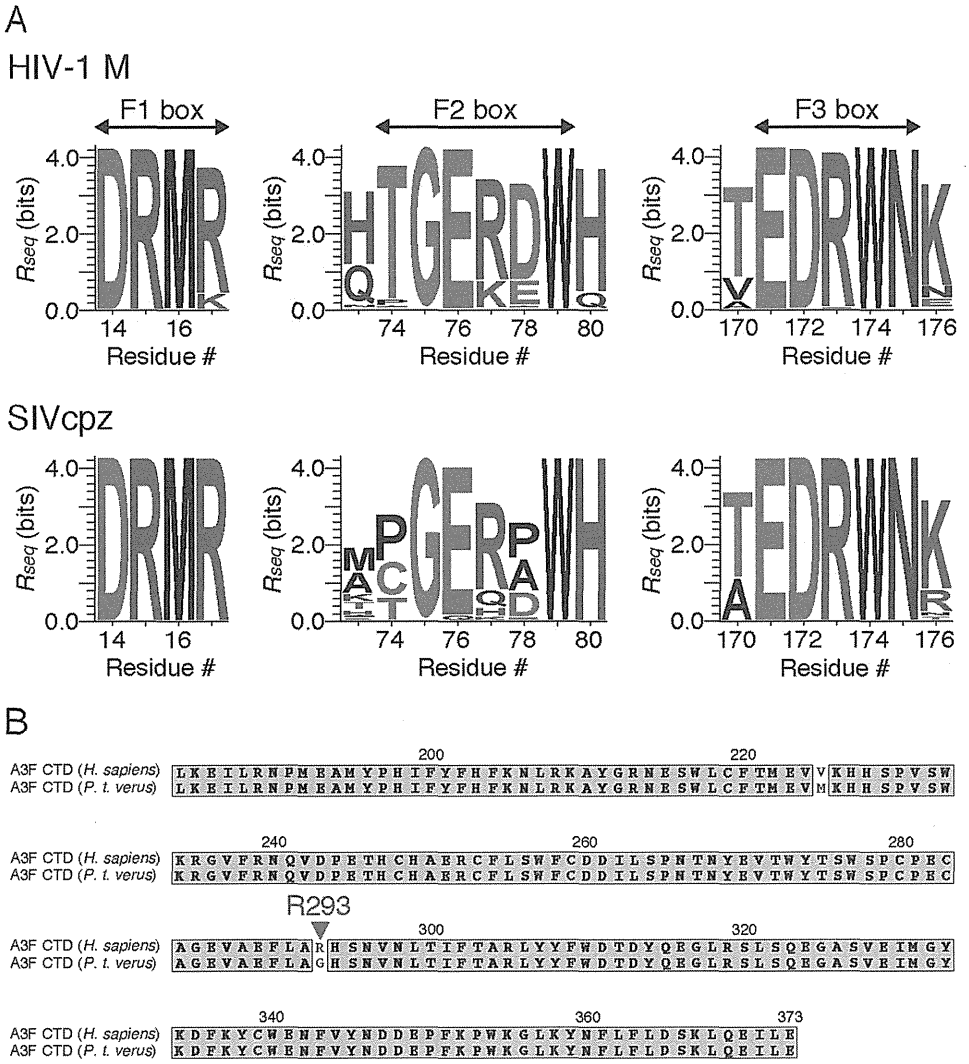


FIG 9 Amino acid sequence conservation of A3F-binding Vif motifs and Vif-binding A3F residues. (A) Conservation at each alignment position in the F1, F2, and F3 boxes. A total of 2,935 HIV-1 sequences and 27 SIVcpz Vif sequences were aligned. The conservation rate at each position (residue) was analyzed using the WebLogo 3.4 program and is shown as R_{seq} values on the *y* axis. Amino acids are colored according to chemical properties: polar (green), neutral (purple), basic (blue), acidic (red), and hydrophobic (black). (B) Amino acid sequence alignment of human (*Homo sapiens*) and chimpanzee (*Pan troglodytes verus*) A3F CTD. Identical residues are shaded in gray. A red arrowhead indicates residue R293 of human A3F.

evolution of the HIV-1/SIVcpz Vif capacity for antagonism, as it was reported that A3D gained increased antiviral activity in the chimpanzee-bonobo lineage (66).

Bohn et al. reported the crystal structure of the A3F CTD with 11 amino acid substitutions and predicted a large continuous surface to be required for the Vif interaction, including 11 critical residues that were previously described (33) and 13 additional residues (58). The prediction was based on their analysis encompassing the negatively charged surface of the A3F CTD structure (58). Indeed, within these 13 residues, we verified that 3 residues (L291, A292, and R293) are critical for the Vif interaction. Hence, including the E324 position (33, 34), four A3F residues are uniquely involved in Vif binding that do not represent equivalent roles in A3C. The previously reported A3F motif EFLARH (residues 289 to 294) includes L291, A292, and R293 (67); the interaction of these residues with Vif was initially missed because our previous investigation was based on reported A3C interface residues (33). A recent study of the HIV-1 Vif sensitivity of rhesus macaque A3F strongly supports that A3F E324 is a critical determinant for the Vif interaction (68). MD simulations of the A3F CTD structure have previously suggested that conformational differences in the α3-α4 bottoms, where the four critical residues are located, might be responsible for differential Vif sensitivity between the E141K A3C and E324K A3F mutants (33). In fact, the α3 positions are in a slightly different position between the A3C crystal structure (PDB ID no. 3VOW, chain A) and the A3F CTD determined in this study (chain A), although the positioning of the α2 and α4 helices is similar (Fig. 7A). The α3 residues are shifted toward α4 in the A3F CTD compared with A3C. This shifted positioning causes a narrower α3-α4 space in the A3F CTD than A3C: the distance between the Cαs of A3F CTD α2 (C259)-α4 (E324) is 0.7 Å shorter than in A3C α2 (C76)-α4 (E141), whereas that of A3F CTD α3 (A292)-α4 (E324) is 2.8 Å shorter than in A3C α3 (A108)-α4 (E141). These results suggest

that A3F positions L291, A292, R293, and E324 may be more sensitive to mutation than A3C, possibly due to induction of local conformational distortion. Interestingly, the A3F R293A or A3F R293G mutants showed a more Vif-resistant phenotype than R293E or R293D, whereas the A292E mutation attenuated Vif sensitivity and Vif-binding ability (Fig. 3A and 4). These results indicate that the presence of a negatively charged region on the A3F surface is not simply associated with specific Vif binding, but rather the shape and distribution of electric charges at the local A3F surface might be critical for the A3F-Vif interaction. Thereby, mutational disruption of such a charge network at the A3F CTD α3-α4 bottom could be vulnerable to Vif-binding ability.

In this work, A3F R293 was identified as one of the most critical residues for the interaction with HIV-1 Vif. Comparative analysis of A3F amino acid sequences among the family Hominidae showed that the chimpanzee carries an A3F gene with an HIV-1 Vif-resistant genotype, G293 instead of R293 (GenBank Accession DQ373064) (Fig. 9). The chimpanzee G293 genotype is also found in the sequence database of the Great Ape Genome Project (69). Adaptation of SIV in chimpanzees cleared cross-species barriers of host restriction factors and became the source of HIV-1 M, which is currently causing a pandemic in humans (70, 71). Indeed, of the four chimpanzee subspecies, *Pan troglodytes troglodytes* (*P. t. troglodytes*) and *P. t. schweinfurthii* have been shown to harbor natural infections of SIVcpz in Central and Eastern Africa, whereas *P. t. verus* does not (71). Additionally, phylogenetic analyses of HIV-1 M and SIVcpz show high genetic similarity (71). The amino acid sequences of Vif are also highly similar between HIV-1 M and SIVcpz, with 91% similarity between HIV-1 HXB2 and SIVcpz LB7 strains and high levels of conservation of the F-box motifs (Fig. 9A). These results prompt several questions in terms of the roles of A3F restriction. (i) Do HIV-1 M and/or SIVcpz Vifs antagonize the antiviral A3F function in chimpanzees? (ii) Does A3F exert only a negligible effect on the restriction of SIVcpz in chimpanzees? (iii) Are there any wild-living chimpanzees with the A3F R293 genotype, especially among naturally infected species in Central and Eastern Africa? Additional investigations are required to answer these questions.

In summary, our structure-based analyses of the A3C/F-binding interfaces on Vif revealed that three discontinuous F-box motifs are assembled at one region of the molecule with hydrophobic and positively charged surface patches. The positively charged Vif stretch contributes to formation of an extended interface critical for the A3F interaction, although it is not essential for the A3C binding. Moreover, our determination of the WT A3F CTD crystal structure and identification of four additional A3F residues critical for the Vif interaction demonstrated that the Vif-binding A3F interface includes an additional negatively charged region, which electrostatically intercorrelates with the positively charged extension on the Vif interface. These findings establish a framework for understanding the structural mechanism of specific recognition between A3F and Vif. Furthermore, these results will advance our understanding of host-virus interactions and antagonisms in terms of A3-Vif during the cross-species transmission that led to the current human pandemic of HIV-1.

ACKNOWLEDGMENTS

We thank Atsuo Suzuki (Nagoya University) for structure data analysis. The TZM-bl cells were provided by J. C. Kappes and X. Wu through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health, NIH Reagent Program.

This work was supported in part by a grant-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by a grant for HIV/AIDS Research from the Ministry of Health, Labor, and Welfare of Japan (to Y.I.) and by a grant-in-aid for the Japan Society for the Promotion of Science (JSPS) Fellows (to M. Nakashima). This work was performed under the approval of the Photon Factory Program Advisory Committee (proposal no. 2012G642 and 2014G559).

The authors declare that no conflict of interest exists.

FUNDING INFORMATION

This work was funded by the Japan Society for the Promotion of Science (JSPS) under grant 15J12567 (Masaaki Nakashima). This work was funded by JSPS under KAKENHI grant 15H04740 (Yasumasa Iwatani) and by the Japan Agency for Medical Research and Development (AMED) (Yasumasa Iwatani). This work was performed under the approval of the Photon Factory Program Advisory Committee (proposal no. 2012G642, 2014G559).

REFERENCES

1. **Albin JS, Harris RS.** 2010. Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med **12**:e4. http://dx.doi.org/10.1017/S1462399409001343.
2. **Desimmie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK.** 2014. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. J Mol Biol **426**:1220–1245. http://dx.doi.org/10.1016/j.jmb.2013.10.033.
3. **Goila-Gaur R, Strebel K.** 2008. HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology **5**:51. http://dx.doi.org/10.1186/1742-4690-5-51.
4. **Malim MH, Emerman M.** 2008. HIV-1 accessory proteins—ensuring viral survival in a hostile environment. Cell Host Microbe **3**:388–398. http://dx.doi.org/10.1016/j.chom.2008.04.008.
5. **LaRue RS, Andresdottir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jonsson SR, Landau NR, Lochelt M, Malik HS, Malim MH, Munk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu XF, Yuhki N, Harris RS.** 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol **83**:494–497. http://dx.doi.org/10.1128/JVI.01976-08.
6. **Smith HC, Bennett RP, Kizilyer A, McDougall WM, Prohaska KM.** 2012. Functions and regulation of the APOBEC family of proteins. Semin Cell Dev Biol **23**:258–268. http://dx.doi.org/10.1016/j.semcdb.2011.10.004.
7. **Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng YH.** 2008. Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem **283**:11606–11614. http://dx.doi.org/10.1074/jbc.M707586200.
8. **Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, Hecht FM, Barbour JD, Simon V.** 2013. HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell Host Microbe **14**:411–421. http://dx.doi.org/10.1016/j.chom.2013.09.006.
9. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418**:646–650. http://dx.doi.org/10.1038/nature00939.
10. **Wiegand HL, Doehle BP, Bogerd HP, Cullen BR.** 2004. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J **23**:2451–2458. http://dx.doi.org/10.1038/sj.emboj.7600246.
11. **Zheng YH, Irwin D, Kurosu T, Tokunaga K, Sata T, Peterlin BM.** 2004. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J Virol **78**:6073–6076. http://dx.doi.org/10.1128/JVI.78.11.6073-6076.2004.
12. **Iwatani Y, Takeuchi H, Strebel K, Levin JG.** 2006. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. J Virol **80**:5992–6002. http://dx.doi.org/10.1128/JVI.02680-05.
13. **Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG.** 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res **35**:7096–7108. http://dx.doi.org/10.1093/nar/gkm750.

14. **Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH.** 2003. DNA deamination mediates innate immunity to retroviral infection. Cell **113**:803–809. http://dx.doi.org/10.1016/S0092-8674(03)00423-9.
15. **Lecossier D, Bouchonnet F, Clavel F, Hance AJ.** 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science **300**:1112. http://dx.doi.org/10.1126/science.1083338.
16. **Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature **424**:99–103. http://dx.doi.org/10.1038/nature01709.
17. **Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L.** 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature **424**:94–98. http://dx.doi.org/10.1038/nature01707.
18. **Bishop KN, Holmes RK, Malim MH.** 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol **80**:8450–8458. http://dx.doi.org/10.1128/JVI.00839-06.
19. **Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM.** 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol **15**:166–170. http://dx.doi.org/10.1016/j.cub.2004.12.068.
20. **Chaurasiya KR, McCauley MJ, Wang W, Qualley DF, Wu T, Kitamura S, Geertsema H, Chan DS, Hertz A, Iwatani Y, Levin JG, Musier-Forsyth K, Rouzina I, Williams MC.** 2014. Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein. Nat Chem **6**:28–33. http://dx.doi.org/10.1038/nchem.1795.
21. **Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF.** 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science **302**:1056–1060. http://dx.doi.org/10.1126/science.1089591.
22. **Jager S, Kim DY, Hultquist JF, Shindo K, LaRue RS, Kwon E, Li M, Anderson BD, Yen L, Stanley D, Mahon C, Kane J, Franks-Skiba K, Cimermancic P, Burlingame A, Sali A, Craik CS, Harris RS, Gross JD, Krogan NJ.** 2012. Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. Nature **481**:371–375. http://dx.doi.org/10.1038/nature10693.
23. **Zhang W, Du J, Evans SL, Yu Y, Yu XF.** 2012. T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. Nature **481**:376–379. http://dx.doi.org/10.1038/nature10718.
24. **Kitamura S, Ode H, Iwatani Y.** 2011. Structural features of antiviral APOBEC3 proteins are linked to their functional activities. Front Microbiol **2**:258. http://dx.doi.org/10.3389/fmicb.2011.00258.
25. **Aydin H, Taylor MW, Lee JE.** 2014. Structure-guided analysis of the human APOBEC3-HIV restrictome. Structure **22**:668–684. http://dx.doi.org/10.1016/j.str.2014.02.011.
26. **Oberste MS, Gonda MA.** 1992. Conservation of amino-acid sequence motifs in lentivirus Vif proteins. Virus Genes **6**:95–102. http://dx.doi.org/10.1007/BF01703760.
27. **Tian C, Yu X, Zhang W, Wang T, Xu R, Yu XF.** 2006. Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F. J Virol **80**:3112–3115. http://dx.doi.org/10.1128/JVI.80.6.3112-3115.2006.
28. **Schrofelbauer B, Senger T, Manning G, Landau NR.** 2006. Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with nonhuman primate APOBEC3G. J Virol **80**:5984–5991. http://dx.doi.org/10.1128/JVI.00388-06.
29. **Russell RA, Pathak VK.** 2007. Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. J Virol **81**:8201–8210. http://dx.doi.org/10.1128/JVI.00395-07.
30. **He Z, Zhang W, Chen G, Xu R, Yu XF.** 2008. Characterization of conserved motifs in HIV-1 Vif required for APOBEC3G and APOBEC3F interaction. J Mol Biol **381**:1000–1011. http://dx.doi.org/10.1016/j.jmb.2008.06.061.
31. **Donahue JP, Vetter ML, Mukhtar NA, D'Aquila RT.** 2008. The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation. Virology **377**:49–53. http://dx.doi.org/10.1016/j.virol.2008.04.017.
32. **Dang Y, Davis RW, York IA, Zheng YH.** 2010. Identification of 81LGxGxxIxW89 and 171EDRW174 domains from human immunodeficiency virus type 1 Vif that regulate APOBEC3G and APOBEC3F neutralizing activity. J Virol **84**:5741–5750. http://dx.doi.org/10.1128/JVI.00079-10.
33. **Kitamura S, Ode H, Nakashima M, Imahashi M, Naganawa Y, Kurosawa T, Yokomaku Y, Yamane T, Watanabe N, Suzuki A, Sugiura W, Iwatani Y.** 2012. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nat Struct Mol Biol **19**:1005–1010. http://dx.doi.org/10.1038/nsmb.2378.
34. **Albin JS, LaRue RS, Weaver JA, Brown WL, Shindo K, Harjes E, Matsuo H, Harris RS.** 2010. A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. J Biol Chem **285**:40785–40792. http://dx.doi.org/10.1074/jbc.M110.173161.
35. **Nguyen K-L, Llano M, Akari H, Miyagi E, Poeschla EM, Strebel K, Bour S.** 2004. Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression. Virology **319**:163–175. http://dx.doi.org/10.1016/j.virol.2003.11.021.
36. **Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol **59**:284–291.
37. **Kao S, Khan MA, Miyagi E, Plishka R, Buckler-White A, Strebel K.** 2003. The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. J Virol **77**:11398–11407. http://dx.doi.org/10.1128/JVI.77.21.11398-11407.2003.
38. **Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, Saag MS, Wu X, Shaw GM, Kappes JC.** 2002. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob Agents Chemother **46**:1896–1905. http://dx.doi.org/10.1128/AAC.46.6.1896-1905.2002.
39. **Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, Kappes JC, Shaw GM, Hunter E.** 2000. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. J Virol **74**:8358–8367. http://dx.doi.org/10.1128/JVI.74.18.8358-8367.2000.
40. **Takeuchi Y, McClure MO, Pizzato M.** 2008. Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. J Virol **82**:12585–12588. http://dx.doi.org/10.1128/JVI.01726-08.
41. **Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D.** 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol **72**:2855–2864.
42. **Platt EJ, Bilska M, Kozak SL, Kabat D, Montefiori DC.** 2009. Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. J Virol **83**:8289–8292. http://dx.doi.org/10.1128/JVI.00709-09.
43. **Iwatani Y, Chan DS, Liu L, Yoshii H, Shibata J, Yamamoto N, Levin JG, Gronenborn AM, Sugiura W.** 2009. HIV-1 Vif-mediated ubiquitination/degradation of APOBEC3G involves four critical lysine residues in its C-terminal domain. Proc Natl Acad Sci U S A **106**:19539–19544. http://dx.doi.org/10.1073/pnas.0906652106.
44. **Otwinowski Z, Minor W.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol **276**:307–326. http://dx.doi.org/10.1016/S0076-6879(97)76066-X.
45. **McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.** 2007. Phaser crystallographic software. J Appl Crystallogr **40**:658–674. http://dx.doi.org/10.1107/S0021889807021206.
46. **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS.** 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr **67**:235–242. http://dx.doi.org/10.1107/S0907444910045749.
47. **Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN.** 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr D Biol Crystallogr **60**:2184–2195. http://dx.doi.org/10.1107/S0907444904023510.
48. **Emsley P, Cowtan K.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr **60**:2126–2132. http://dx.doi.org/10.1107/S0907444904019158.
49. **Schrodinger LLC.** 2010. The PyMOL molecular graphics system, version 1.6. Schrodinger, LLC, New York, NY.

50. **Guo Y, Dong L, Qiu X, Wang Y, Zhang B, Liu H, Yu Y, Zang Y, Yang M, Huang Z.** 2014. Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. Nature **505**:229–233. http://dx.doi.org/10.1038/nature12884.
51. **Sali A, Blundell TL.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234**:779–815. http://dx.doi.org/10.1006/jmbi.1993.1626.
52. **Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA.** 2006. AMBER 9. University of California, San Francisco, CA.
53. **Pang YP, Xu K, Yazal JE, Prendergas FG.** 2000. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. Protein Sci **9**:1857–1865.
54. **Schueler-Furman O, Wang C, Baker D.** 2005. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. Proteins **60**:187–194. http://dx.doi.org/10.1002/prot.20556.
55. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797. http://dx.doi.org/10.1093/nar/gkh340.
56. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res **14**:1188–1190. http://dx.doi.org/10.1101/gr.849004.
57. **Simon JH, Southerling TE, Peterson JC, Meyer BE, Malim MH.** 1995. Complementation of vif-defective human immunodeficiency virus type 1 by primate, but not nonprimate, lentivirus vif genes. J Virol **69**:4166–4172.
58. **Bohn MF, Shandilya SM, Albin JS, Kouno T, Anderson BD, McDougle RM, Carpenter MA, Rathore A, Evans L, Davis AN, Zhang J, Lu Y, Somasundaran M, Matsuo H, Harris RS, Schiffer CA.** 2013. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. Structure **21**:1042–1050. http://dx.doi.org/10.1016/j.str.2013.04.010.
59. **Siu KK, Sultana A, Azimi FC, Lee JE.** 2013. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. Nat Commun **4**:2593. http://dx.doi.org/10.1038/ncomms3593.
60. **Holmes RK, Koning FA, Bishop KN, Malim MH.** 2007. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. J Biol Chem **282**:2587–2595.
61. **Byeon IJ, Ahn J, Mitra M, Byeon CH, Hercik K, Hritz J, Charlton LM, Levin JG, Gronenborn AM.** 2013. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. Nat Commun **4**:1890. http://dx.doi.org/10.1038/ncomms2883.
62. **Shandilya SM, Nalam MN, Nalivaika EA, Gross PJ, Valesano JC, Shindo K, Li M, Munson M, Royer WE, Harjes E, Kono T, Matsuo H, Harris RS, Somasundaran M, Schiffer CA.** 2010. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure **18**:28–38. http://dx.doi.org/10.1016/j.str.2009.10.016.
63. **Etienne L, Hahn BH, Sharp PM, Matsen FA, Emerman M.** 2013. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. Cell Host Microbe **14**:85–92. http://dx.doi.org/10.1016/j.chom.2013.06.002.
64. **Yu Q, Chen D, Konig R, Mariani R, Unutmaz D, Landau NR.** 2004. APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. J Biol Chem **279**:53379–53386. http://dx.doi.org/10.1074/jbc.M408802200.
65. **Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS.** 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol **85**:11220–11234. http://dx.doi.org/10.1128/JVI.05238-11.
66. **Duggal NK, Malik HS, Emerman M.** 2011. The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. J Virol **85**:11361–11371. http://dx.doi.org/10.1128/JVI.05046-11.
67. **Smith JL, Pathak VK.** 2010. Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. J Virol **84**:12599–12608. http://dx.doi.org/10.1128/JVI.01437-10.
68. **Land AM, Shaban NM, Evans L, Hultquist JF, Albin JS, Harris RS.** 2014. APOBEC3F determinants of HIV-1 Vif sensitivity. J Virol **88**:12923–12927. http://dx.doi.org/10.1128/JVI.02362-14.
69. **Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prufer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprubi C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, et al.** 2013. Great ape genetic diversity and population history. Nature **499**:471–475. http://dx.doi.org/10.1038/nature12228.
70. **Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH.** 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397**:436–441. http://dx.doi.org/10.1038/17130.
71. **Sharp PM, Hahn BH.** 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med **1**:a006841. http://dx.doi.org/10.1101/cshperspect.a006841.

Antiviral Research 119 (2015) 84–88

Contents lists available at ScienceDirect

Antiviral Research

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

Short Communication

Natural polymorphism S119R of HIV-1 integrase enhances primary INSTI resistance ☆

Atsuko Hachiya [a,*], Hirotaka Ode [a], Masakazu Matsuda [a], Yumiko Kito [a], Urara Shigemi [a], Kazuhiro Matsuoka [a], Junji Imamura [a], Yoshiyuki Yokomaku [a], Yasumasa Iwatani [a,b], Wataru Sugiura [a,b]

[a] Department of Infectious Disease and Immunology, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Aichi 460-0001, Japan
[b] Department of AIDS Research, Graduated School of Medicine Nagoya University, Nagoya, Aichi 460-0001, Japan

ARTICLE INFO

Article history:
Received 18 February 2015
Revised 25 April 2015
Accepted 27 April 2015
Available online 5 May 2015

Keywords:
HIV-1
Integrase strand transfer inhibitor
Drug resistance
Polymorphism
Deep sequencing

ABSTRACT

Integrase strand transfer inhibitors (INSTIs), which block proviral DNA integration into the host chromosome, are clinically effective against HIV-1 isolates exhibiting resistance to other classes of antiretroviral agents. Although naturally occurring amino acid variation has been less frequently observed in the integrase region, the functional constraints of this variation on primary INSTI resistance-associated mutations are not fully understood.

In the present study, we focused on the S119G/R/P/T (S119X) polymorphisms, which are frequently observed in HIV-1 sequences derived from clinical specimens (naïve, $n = 458$, 26%). The frequency of the S119X polymorphism together with Q148H/R ($n = 8$, 63%) or N155H ($n = 12$, 83%) was relatively high compared with that of naïve group. Our *in vitro* assays revealed that S119G/P/T alone exerted no effect on the susceptibility to INSTIs, whereas S119R enhanced the level of INSTI resistance induced by well-known INSTI resistance-associated mutations (Y143C, Q148H or N155H). Notably, the S119R polymorphism contributed to a significant (5.9-fold) increase in dolutegravir resistance caused by G140S/Q148H. Analysis of two cases of virological failure during raltegravir-based therapy showed that the accumulation and the rapid evolution of primary INSTI resistance-associated mutations coincided with the S119R mutation. These data highlight the role of the S119X polymorphism in INSTI resistance, and this polymorphism might be linked to the potential treatment outcome with INSTI-based therapy.

Integrase strand transfer inhibitors (INSTIs), which are currently key drugs in the first-line HIV-1 treatment regimen, are clinically superior to other classes of drugs (Messiaen et al., 2013). INSTIs bind to a catalytic site in the catalytic core domain (CCD) of integrase (IN) and block the strand transfer reaction catalyzed by HIV-1-expressed IN (Hazuda et al., 2000; Kobayashi et al., 2011; Shimura et al., 2008). Therefore, primary HIV-1 mutations conferring INSTI resistance (e.g., G140S, Q148H and N155H) structurally map near the IN catalytic center (Hare et al., 2010; Li et al., 2011; Menendez-Arias, 2013). The polymorphism rate of the CCD of IN variants derived from INSTI-naïve patients has been shown to be low (34%) (Rhee et al., 2008), suggesting that IN sequence conservation may be associated with structural and functional constraints. However, how such constraints affect the appearance of INSTI resistance-associated mutations in the presence of certain polymorphisms is unknown. Here, to investigate the mechanisms underlying the relationship between certain polymorphisms and INSTI resistance-associated mutations, we analyzed the HIV-1 IN polymorphism frequency in clinical specimens.

HIV-1 subtype B IN polymorphism frequencies (%) were evaluated using specimens from HIV-1-infected patients at the National Hospital Organization Nagoya Medical Center, Japan (Fig. 1). The Institutional Ethical Committee approved this study (2010–310), and informed consent was obtained from each individual. A total of 600 IN sequences were analyzed; 458 samples were obtained from antiretroviral therapy (ART)-naïve individuals, and 142 samples were obtained from ART-experienced individuals, including patients carrying the Y143C ($n = 1$), Q148H/R ($n = 8$) or N155H ($n = 12$) mutation. The frequency of HIV-1 integrase polymorphisms analyzed in our Japanese cohorts was similar to previous data from the Italian Cohort of Antiretroviral Naïve (ICONA)

Abbreviations: HIV-1, human immunodeficiency virus type-1; INSTI, integrase strand transfer inhibitor; ART, antiretroviral therapy; CCD, catalytic core domain.

☆ *Database linking:* Direct sequence and deep sequence data reported in this paper have been deposited in the GenBank (Accession Nos.: LC022131–LC022730, and Accession No.: DRA003039).

* Corresponding author at: Clinical Research Center, National Hospital Organization Nagoya Medical Center, 4-1-1 San-no-maru, Naka-ku, Nagoya, Aichi 460-0001, Japan. Tel.: +81 52 951 1111; fax: +81 52 963 3970.
E-mail address: hachiya@nnh.hosp.go.jp (A. Hachiya).

http://dx.doi.org/10.1016/j.antiviral.2015.04.014
0166-3542/

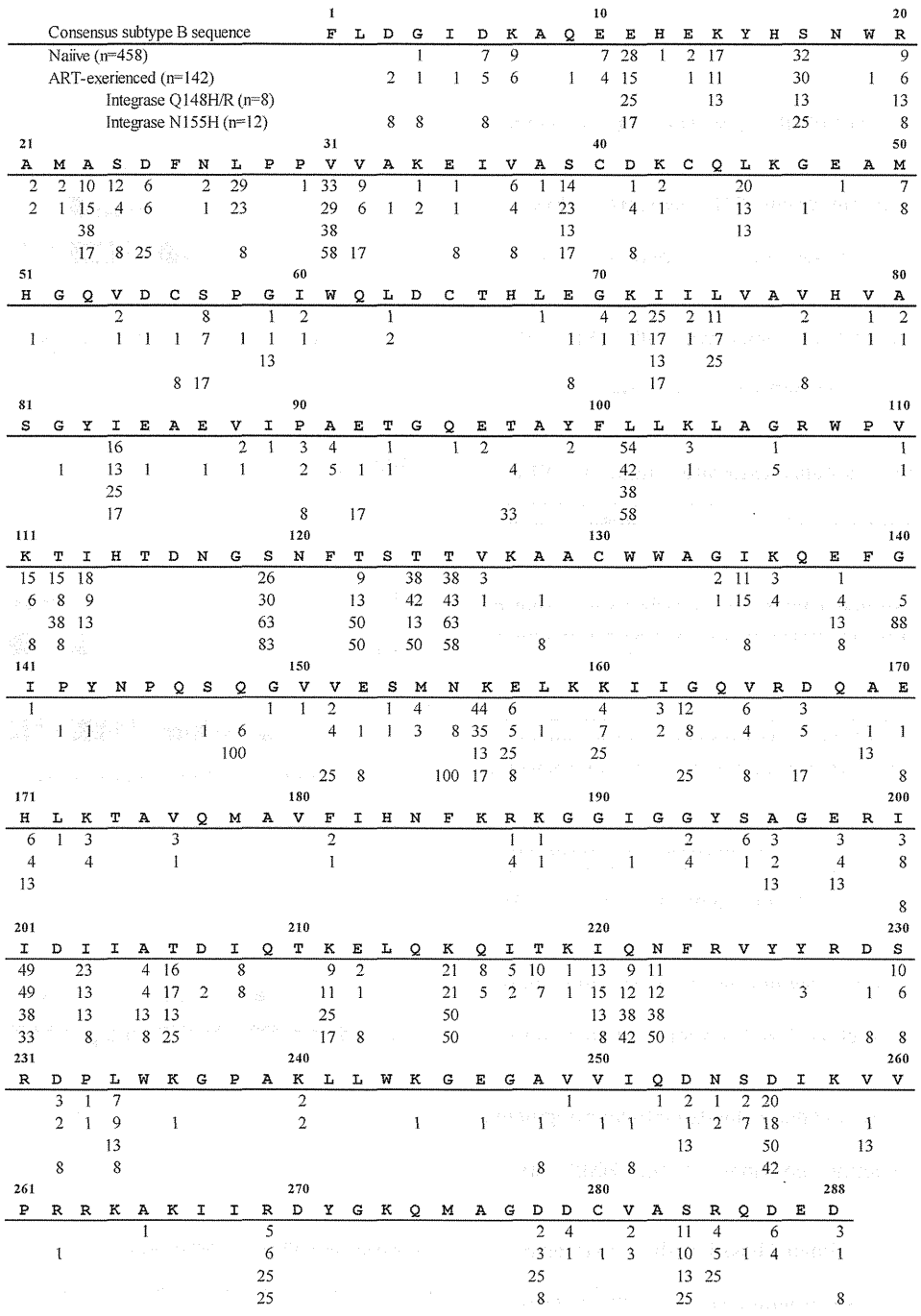


Fig. 1. Frequency of polymorphisms in HIV-1 subtype B IN sequences derived from clinical specimens. Consensus sequences of IN from HIV-1 subtype B are shown at the top. The polymorphism frequency (%) in the HIV-1 IN region among ART-naïve and ART-experienced individuals infected with HIV-1 are shown (GenBank Accession Nos: LC022131 to LC022730). HIV-1 carrying Q148H/R or N155H were classified as a subgroup of ART-experienced. The frequency values are presented only for mutations displaying a frequency >0.5%.

(Ceccherini-Silberstein et al., 2010). Besides E11, S17, I72, L101, S119, T124 and T125, we observed integrase polymorphisms (>25%) at L28, V31, K156 and I201 in HIV-1 (subtype B) derived from the ART-Naïve patients. Especially, the positions at E11, S17, V31, I72, L101, S119, T124, T125 and I201 are also highly polymorphic among the other subtypes based on two major online repositories; the Stanford HIV-1 Databases and the Los Alamos National Laboratory (Vavro et al., 2013). We found no significant accumulation of previously reported minor INSTI resistance-associated mutations at Q95, H183, Y226, S230, or D232 (Rhee et al., 2008). Interestingly, in the IN sequences derived from all samples, multiple amino acid patterns were observed at position 119: S119T (10.7%), S119P (8.2%), S119G (7.6%), and S119R (1.2%). Additionally, the S119G/R/P/T (S119X) polymorphism frequency was significantly higher in the ART-experienced group, which specifically carried Q148H/R or N155H (63% and 83%, respectively), than in the ART-naïve group (26%). HIV-1 carrying Y143C ($n = 1$) also carried S119R. Furthermore, according to the IN sequences deposited in the HIV Drug Resistance Database of Stanford University (http://hivdb.stanford.edu/), S119X polymorphisms are more frequently observed in individuals infected with HIV-1 carrying N155H (41.9%, $n = 93$) than in individuals with

N155 (33.2%, $n = 4302$). These data suggest that the primary INSTI resistance-associated mutations tend to accompany S119X polymorphisms.

Next, to investigate whether S119X polymorphisms are linked to the level of INSTI resistance of HIV-1, we generated 4 recombinant viruses carrying G, P, T or R substitutions at position 119 of IN as previously described (Shimura et al., 2008). We tested the resistance level of each virus to raltegravir (RAL), elvitegravir (EVG), or dolutegravir (DTG) using a TZM-bl cell-based assay system (Hachiya et al., 2013). As shown in Fig. 2 and Supplemental Tables 1–3, HIV-1 isolates carrying S119G, S119P, or S119T exhibited slightly enhanced (~0.6-fold) susceptibility to RAL and similar susceptibility to EVG and DTG compared with HIV-1 S119 (WT). However, HIV-1 S119R showed a tendency toward slightly increased resistance to all three INSTIs (1.8-, 1.4- and 1.7-fold for RAL, EVG and DTG, respectively, compared with WT).

Moreover, to clarify the impact of adding S119X polymorphisms to the well-known INSTI resistance-associated mutations Y143C, Q148H, N155H, G140S/Q148H or T97A/Y143C on INSTI resistance, we constructed 21 recombinant viruses carrying different mutation combinations and compared their INSTI resistance levels. The results showed that INSTI-resistant viruses carrying S119G/P/T exhibited a similar level of INSTI resistance to that of the S119 virus, indicating no effect of the S119G/P/T polymorphisms on INSTI resistance. By contrast, S119R significantly increased the RAL resistance of HIV-1 Y143C, Q148H, N155H, G140S/Q148H or T97A/Y143C (4.9-, 26-, 16-, 479- and 27-fold, respectively, compared with WT). Additionally, S119R enhanced the EVG resistance of HIV-1 Q148H and N155H (12- and 95-fold, respectively, compared with WT) but not HIV-1 Y143C, consistent with the fact that Y143C is a primary resistance mutation for RAL, but not EVG (Wensing et al., 2014). Furthermore, the S119R polymorphism enhanced the resistance of HIV-1 G140S/Q148H (5.9-fold) to DTG. These *in vitro* data demonstrate that the S119R polymorphism increases the resistance of primary INSTI-resistant HIV-1 to INSTIs.

To further assess the clinical significance of these results in the context of S119R and the INSTI resistance-associated mutations, we analyzed two cases of clinical and virological failure during RAL-based therapy using the clinical histories and the deep sequencing-based HIV-1 genotyping results of two individuals (Fig. 3). Near full-length HIV-1 genome divided into four fragments (nucleotides 705–3402, 2509–5192, 4168–7653 and 6895–9531 according to the numbering positions of the HXBII strain, GenBank ID# K03455) were amplified and sequenced using the MiSeq platform (Illumina) (Ode et al., 2014). The within-patients mutation frequencies were estimated based on the generated reads using VICUNA (Yang et al., 2012), Burrows–Wheeler Aligner (Li and Durbin, 2009) and our in-house program (Ode et al., 2014). For case 1 (Fig. 3A and C), HIV-1 genotypic assays were performed prior to RAL-based therapy (baseline) and at three different time points in which the HIV-1 viral load (VL) rebounded. At baseline (#1, August 2005), during tenofovir disoproxil fumarate (TDF)/lamivudine (3TC)/efavirenz (EFV) treatment, no primary INSTI resistance-associated mutations were detected. However, when the HIV-1 VL rebounded to 1000 copies/ml after switching to maraviroc (MVC)/RAL treatment, a primary RAL resistance mutation, N155H, appeared as a major variant (~99.9% of the total viral population) between time points #2 and #4 (Fig. 3C). Minor viral populations (less than 20%) carrying other polymorphisms, such as K14N, G47E G70E, L74M, I84L, G163R, S230R, and I251L, were not increased between time points #3 and #4. However, the S119R frequency was drastically increased (from 38.9% to 90%) with the increasing frequency of N155H among the HIV-1 population. Notably, we confirmed the coexistence of S119R with N155H in the same sequence reads.

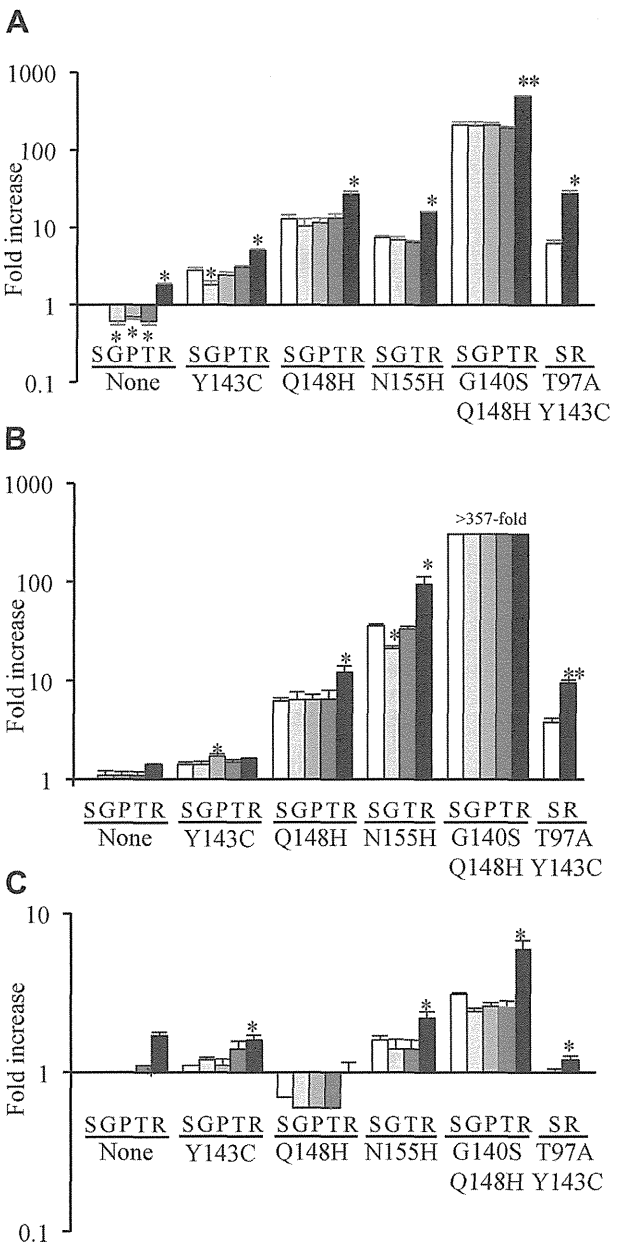
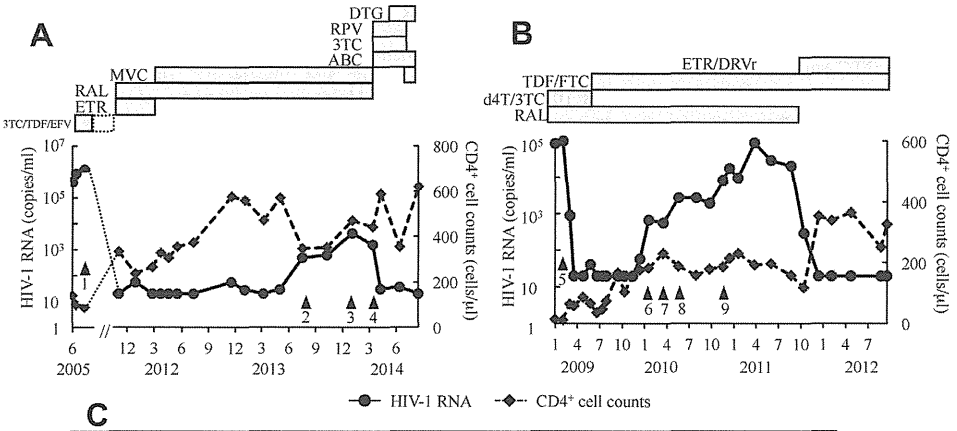


Fig. 2. Susceptibility of HIV isolates carrying the S119X polymorphism in the background of well-known INSTI resistance-associated mutations to INSTIs. The antiviral activities of RAL (A), EVG (B) or DTG (C) in HIV-1 isolates carrying a mutation at residue 119 (S119, S119G, S119P, S119T or S119R) in the WT (none) or well-known INSTI resistance-associated mutation background (Y143C, Q148H, N155H, G140S/Q148H or T97A/Y143C) were determined using the TZM-bl assay system. The fold increase is expressed as the EC_{50} value for the recombinant viruses compared with the WT virus (S119/none). The error bars represent the standard deviation from three independent experiments. The asterisks indicate significant differences in the EC_{50} value ($^{*}p < 0.05$ and $^{**}p < 0.0001$ based on the *t*-test). The susceptibilities of HIV-1 Q140S/Q148H and S119X to EVG were not assessed because their EC_{50} values were above the detection limit for this assay (>1 μM). The susceptibility of HIV-1 S119P/N155H was not tested.

In case 2, the HIV-1 VL rebounded to 100,000 copies/ml during TDF/emtricitabine (FTC)/RAL treatment (Fig. 3B and D). As shown in Fig. 3D, the secondary RAL resistance-associated mutation T97A appeared simultaneously with S119R at time point #7 and increased between time points #7 and #8 (from 42.4% to 100%) in proportions similar to those of S119R (from 42.4% to 99.9%), suggesting the coincident appearance of S119R and T97A. At time points #8 and #9, Y143C mutation in the T97A/S119R variant



C

Time point	K14N	G47E	G70E	L74M	I84L	**S119R**	**N155H**	G163R	I204V	S230R	I251L
1	-	-	-	-	-	-	-	-	43.7	-	-
2	-	-	-	-	-	-	**99.9**	-	-	-	-
3	4.3	-	3.3	4.3	3.1	**38.9**	**100**	9.0	-	20	17.1
4	-	16.3	-	-	-	**90**	**100**	-	-	17	17

D

Time point	K14R	A21T	L63I	**T97A**	**S119R**	A124T	**Y143X**	K173R	V176L	F181L	G189R	N232D	S255N	K259I
5	-	-	7.5	-	-	15.7	-	4.6	41.1	57.9	-	100	-	-
6	-	-	-	-	-	-	-	-	-	100	-	100	-	-
7	-	46.3	-	**42.4**	**42.4**	-	-	-	-	99.7	-	100	-	-
8	-	-	-	**100**	**99.9**	-	**80.4 (C)** 19.1 (S)	-	-	99.8	5.1	-	4.9	-
9	6.5	-	-	**99.2**	**99.9**	-	**93.5 (C)** 4.7 (A)	-	-	99.6	-	-	-	56.1

Fig. 3. Clinical course and drug resistance profile of two patients exhibiting RAL treatment failure. (A and B) The treatment histories and clinical courses of case 1 (A) and 2 (B). The triangles indicate the time points for deep sequencing-based HIV-1 genotyping assays (GenBank Accession No. DRA003039). The virologic responses represented by plasma HIV-1 viral load (solid line with circle) and CD4$^+$ T lymphocyte counts (dotted line with diamond). Abbreviations of drugs used: FTC, emtricitabine; d4T, stavudine; 3TC, lamivudine; ABC, abacavir; TDF, tenofovir; EFV, efavirenz; ETR, etravirine; RPV, rilpivirine; RAL, raltegravir; DTG, dolutegravir; DRVr, ritonavir-boosted darunavir; MVC, maraviroc. (A) None of the nucleotide reverse transcriptase inhibitor (NRTI) and/or protease inhibitor (PI) associated INSTI resistance mutations was observed at baseline (time point #1). CCR5 (R5) tropic virus was dominant based on the HIV-1 coreceptor tropism assay (geno2pheno, http://coreceptor.geno2pheno.org/) at time point #4. (B) The minor PI resistance mutation M46I/L and the NRTI revertant mutation T215S were observed at baseline (time point #5). (C and D) The HIV-1 genotyping data analyzed by deep-sequencing in panels C and D correspond to the two cases shown in panels A and B, respectively. The frequencies (%) of the INSTI resistance-associated mutations were analyzed using viral RNA from plasma. Briefly, near full-length HIV-1 genome divided into four fragments were amplified, purified and validated by Qubit Fluorometer (Life technologies) and Agilent 2100 Bioanalyzer (Agilent Technologies). PCR amplicons were subjected to tagmentation and to DNA denaturation, and were sequenced on the Illumina MiSeq (paired end 250 bp sequencing read). A minimum coverage was 1000 per nucleotide position ensured to identify a minor variant present. The INSTI resistance-associated mutations reported by the International AIDS Society-USA (Wensing et al., 2014) are shown in bold. The frequency (%) of each mutation is represented only for the mutations displaying a frequency >3%.

was observed in the major viral population. Therefore, S119R polymorphism might be linked to the treatment outcome for RAL-treated patients, possibly leading to virological failure.

Recently, Brockman et al. identified an uncommon HLA-C*05-restricted epitope ($HTDNGSNF_{114-121}$) in the IN region that likely contributes to S119X polymorphism selection; S119R was found at a frequency of 13.9% in C*05-expressing ART-naïve individuals, whereas S119P was observed in 40.2% of these individuals (Brockman et al., 2012). In the present study, S119R, together with INSTI resistance-associated mutations, was selected during RAL-containing therapy in two patients not carrying HLA C*05. Therefore, the S119R polymorphism may be selected by INSTI treatment pressure as well as by HLA-specific host immunity.

Residue 119 substitutions alter integration target specificity (Harper et al., 2001). Recent reports using deep sequencing methods have shown that HIV-1 isolates carrying S119G favor integration sites far from gene-dense chromatin regions but that viruses carrying S119R exhibit global integration profiles compared with S119-carrying viruses (Demeulemeester et al., 2014). At the sites of strand transfer, S119R showed preference A:T bp at positions -2, -3 (and 6, 7) of the viral integration site that are distinct from base distributions by S119 (Demeulemeester et al., 2014). Furthermore, polymorphism in prototype foamy virus (PFV) IN A188 (a residue structurally equivalent to HIV-1 S119) also showed a bias for base distribution, resulting that it altered strand transfer activity (Maertens et al., 2010). Such distinct integration preferences conferred by the S119S polymorphisms might be indirectly linked to primary INSTI resistance mutations and affected strand transfer activity. Notably, our analysis of multi-cycle replication kinetics showed that the S119R mutation further decreased the replication capacity of INSTI-resistant HIV-1 conferred by major resistance mutations, e.g., N155H or Q148H (Supplemental Fig. 1).

Recent biochemical studies have demonstrated that DTG dissociates from the HIV-1 IN-DNA complex more slowly than RAL or EVG; as a result, DTG displays a higher genetic barrier to INSTI resistance (Hightower et al., 2011). The accumulated evidence on the clinical use of INSTIs has shown that DTG effectively suppresses virus replication in individuals (Clotet et al., 2014; Raffi et al., 2013; Walmsley et al., 2013). Moreover, according to *in vitro* assays, DTG retains its inhibitory effect against most

HIV-1 isolates carrying the primary INSTI resistance-associated mutations to RAL and/or EVG (Kobayashi et al., 2011). Even in the presence of two primary DTG resistance-associated mutations, G140S/Q148H, the virus exhibited an only 2.6-fold increase in DTG resistance (Kobayashi et al., 2011). Because our *in vitro* assays demonstrated that S119R contributes to an (5.9-fold) increase in the DTG resistance caused by G140S/Q148H, further analysis of DTG resistance profiles and their molecular mechanisms are required for the development of safe and effective HIV therapy.

In conclusion, we showed that the IN S119R polymorphism, which is frequently observed in clinical specimens, enhances INSTI resistance of INSTI-resistant HIV-1. Thus, this polymorphism might represent a potential risk factor for failure of INSTI-based therapy. These findings provide important information about INSTI resistance profiles and facilitate the further development of novel antiretroviral drugs targeting INSTI-resistant HIV-1.

Acknowledgments

This work was supported by a grant for the promotion of AIDS research from the Ministry of Health, Labor and Welfare (Y.Y., Y.I. and W.S.). We thank Dr. Eiichi Kodama for his generous gift of the plasmids used in this study, Ms. Reiko Okazaki and Masumi Hosaka for the sample preparation and the CARES staff for their dedicated assistance. We are indebted to the AIDS Research Reference and Reagent Program, NIAID, and NIH for the TZM-bl cells.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.antiviral.2015.04.014.

References

Brockman, M.A., Chopera, D.R., Olvera, A., Brumme, C.J., Sela, J., Markle, T.J., Martin, E., Carlson, J.M., Le, A.Q., McGovern, R., Cheung, P.K., Kelleher, A.D., Jessen, H., Markowitz, M., Rosenberg, E., Frahm, N., Sanchez, J., Mallal, S., John, M., Harrigan, P.R., Heckerman, D., Brander, C., Walker, B.D., Brumme, Z.L., 2012. Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. J. Virol. 86, 6913–6923.

Ceccherini-Silberstein, F., Malet, I., Fabeni, L., Dimonte, S., Svicher, V., D'Arrigo, R., Artese, A., Costa, G., Bono, S., Alcaro, S., Monforte, A., Katlama, C., Calvez, V., Antinori, A., Marcelin, A.G., Perno, C.F., 2010. Specific HIV-1 integrase polymorphisms change their prevalence in untreated versus antiretroviral-treated HIV-1-infected patients, all naive to integrase inhibitors. J. Antimicrob. Chemother. 65, 2305–2318.

Clotet, B., Feinberg, J., van Lunzen, J., Khuong-Josses, M.A., Antinori, A., Dumitru, I., Pokrovskiy, V., Fehr, J., Ortiz, R., Saag, M., Harris, J., Brennan, C., Fujiwara, T., Min, S., Team, I.N.G.S., 2014. Once-daily dolutegravir versus darunavir plus ritonavir in antiretroviral-naive adults with HIV-1 infection (FLAMINGO): 48 week results from the randomised open-label phase 3b study. Lancet 383, 2222–2231.

Demeulemeester, J., Vets, S., Schrijvers, R., Madlala, P., De Maeyer, M., De Rijck, J., Ndung'u, T., Debyser, Z., Gijsbers, R., 2014. HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. Cell Host Microbe 16, 651–662.

Hachiya, A., Reeve, A.B., Marchand, B., Michailidis, E., Ong, Y.T., Kirby, K.A., Leslie, M.D., Oka, S., Kodama, E.N., Rohan, L.C., Mitsuya, H., Parniak, M.A., Sarafianos, S.G., 2013. Evaluation of combinations of 4′-ethynyl-2-fluoro-2′-deoxyadenosine with clinically used antiretroviral drugs. Antimicrob. Agents Chemother. 57, 4554–4558.

Hare, S., Vos, A.M., Clayton, R.F., Thuring, J.W., Cummings, M.D., Cherepanov, P., 2010. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc. Natl. Acad. Sci. U.S.A. 107, 20057–20062.

Harper, A.L., Skinner, L.M., Sudol, M., Katzman, M., 2001. Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. J. Virol. 75, 7756–7762.

Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., Miller, M.D., 2000. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 287, 646–650.

Hightower, K.E., Wang, R., Deanda, F., Johns, B.A., Weaver, K., Shen, Y., Tomberlin, G.H., Carter 3rd., H.L., Broderick, T., Sigethy, S., Seki, T., Kobayashi, M., Underwood, M.R., 2011. Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase–DNA complexes. Antimicrob. Agents Chemother. 55, 4552–4559.

Kobayashi, M., Yoshinaga, T., Seki, T., Wakasa-Morimoto, C., Brown, K.W., Ferris, R., Foster, S.A., Hazen, R.J., Miki, S., Suyama-Kagitani, A., Kawauchi-Miki, S., Taishi, T., Kawasuji, T., Johns, B.A., Underwood, M.R., Garvey, E.P., Sato, A., Fujiwara, T., 2011. In Vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. Antimicrob. Agents Chemother. 55, 813–821.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

Li, X., Krishnan, L., Cherepanov, P., Engelman, A., 2011. Structural biology of retroviral DNA integration. Virology 411, 194–205.

Maertens, G.N., Hare, S., Cherepanov, P., 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468, 326–329.

Menendez-Arias, L., 2013. Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments. Antiviral Res. 98, 93–120.

Messiaen, P., Wensing, A.M., Fun, A., Nijhuis, M., Brusselaers, N., Vandekerckhove, L., 2013. Clinical use of HIV integrase inhibitors: a systematic review and meta-analysis. PLoS ONE 8, e52562.

Ode, H., Matsuoka, K., Matsuda, M., Hachiya, A., Hattori, J., Yokomaku, Y., Iwatani, Y., Sugiura, W., 2014. HIV-1 near full-length genome analysis by next-generation sequencing: evaluation of quasispecies and minority drug resistance. In: International Workshop on Antiviral Drug Resistance: Meeting the Global Challenge, Antiviral Therapy, Berlin, Germany, p. A80.

Raffi, F., Rachlis, A., Stellbrink, H.J., Hardy, W.D., Torti, C., Orkin, C., Bloch, M., Podzamczer, D., Pokrovsky, V., Pulido, F., Almond, S., Margolis, D., Brennan, C., Min, S., S-S Group, 2013. Once-daily dolutegravir versus raltegravir in antiretroviral-naive adults with HIV-1 infection: 48 week results from the randomised, double-blind, non-inferiority SPRING-2 study. Lancet 381, 735–743.

Rhee, S.Y., Liu, T.F., Kiuchi, M., Zioni, R., Gifford, R.J., Holmes, S.P., Shafer, R.W., 2008. Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. Retrovirology 5, 74.

Shimura, K., Kodama, E., Sakagami, Y., Matsuzaki, Y., Watanabe, W., Yamataka, K., Watanabe, Y., Ohata, Y., Doi, S., Sato, M., Kano, M., Ikeda, S., Matsuoka, M., 2008. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). J. Virol. 82, 764–774.

Vavro, C., Hasan, S., Madsen, H., Horton, J., DeAnda, F., Martin-Carpenter, L., Sato, A., Cuffe, R., Chen, S., Underwood, M., Nichols, G., 2013. Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). Antimicrob. Agents Chemother. 57, 1379–1384.

Walmsley, S.L., Antela, A., Clumeck, N., Duiculescu, D., Eberhard, A., Gutierrez, F., Hocqueloux, L., Maggiolo, F., Sandkovsky, U., Granier, C., Pappa, K., Wynne, B., Min, S., Nichols, G., Investigators, S., 2013. Dolutegravir plus abacavir–lamivudine for the treatment of HIV-1 infection. N. Engl. J. Med. 369, 1807–1818.

Wensing, A.M., Calvez, V., Gunthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., Shafer, R.W., Richman, D.D., 2014. 2014 Update of the drug resistance mutations in HIV-1. Topics Antivir. Med. 22, 642–650.

Yang, X., Charlebois, P., Gnerre, S., Coole, M.G., Lennon, N.J., Levin, J.Z., Qu, J., Ryan, E.M., Zody, M.C., Henn, M.R., 2012. De novo assembly of highly diverse viral populations. BMC Genomics 13, 475.

frontiers in Microbiology

METHODS
published: 12 November 2015
doi: 10.3389/fmicb.2015.01258

Quasispecies Analyses of the HIV-1 Near-full-length Genome With Illumina MiSeq

Hirotaka Ode[1‡], Masakazu Matsuda[1‡], Kazuhiro Matsuoka[1†‡], Atsuko Hachiya[1], Junko Hattori[1†], Yumiko Kito[1], Yoshiyuki Yokomaku[1], Yasumasa Iwatani[1,2] and Wataru Sugiura[1,2†]*

[1] Department of Infectious Diseases and Immunology, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan, [2] Department of AIDS Research, Graduate School of Medicine, Nagoya University, Nagoya, Japan

OPEN ACCESS

Edited by:
Francois Villinger,
Emory University School of Medicine, USA

Reviewed by:
Kazuhisa Yoshimura,
National Institute of Infectious Diseases, Japan
Siddappa Byrareddy,
Emory University, USA

***Correspondence:**
Hirotaka Ode
odehir@mail-nmc.jp

†Present Address:
Kazuhiro Matsuoka,
Proteo-Science Center, Ehime University, Ehime, Japan;
Junko Hattori,
HIV Dynamics and Replication Program, National Cancer Institute, National Institutes of Health, Frederick, USA;
Wataru Sugiura,
GlaxoSmithKline, Tokyo, Japan

‡These authors have contributed equally to this work.

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 12 September 2015
Accepted: 29 October 2015
Published: 12 November 2015

Citation:
Ode H, Matsuda M, Matsuoka K, Hachiya A, Hattori J, Kito Y, Yokomaku Y, Iwatani Y and Sugiura W (2015) Quasispecies Analyses of the HIV-1 Near-full-length Genome With Illumina MiSeq. Front. Microbiol. 6:1258. doi: 10.3389/fmicb.2015.01258

Human immunodeficiency virus type-1 (HIV-1) exhibits high between-host genetic diversity and within-host heterogeneity, recognized as quasispecies. Because HIV-1 quasispecies fluctuate in terms of multiple factors, such as antiretroviral exposure and host immunity, analyzing the HIV-1 genome is critical for selecting effective antiretroviral therapy and understanding within-host viral coevolution mechanisms. Here, to obtain HIV-1 genome sequence information that includes minority variants, we sought to develop a method for evaluating quasispecies throughout the HIV-1 near-full-length genome using the Illumina MiSeq benchtop deep sequencer. To ensure the reliability of minority mutation detection, we applied an analysis method of sequence read mapping onto a consensus sequence derived from *de novo* assembly followed by iterative mapping and subsequent unique error correction. Deep sequencing analyses of aHIV-1 clone showed that the analysis method reduced erroneous base prevalence below 1% in each sequence position and discarded only <1% of all collected nucleotides, maximizing the usage of the collected genome sequences. Further, we designed primer sets to amplify the HIV-1 near-full-length genome from clinical plasma samples. Deep sequencing of 92 samples in combination with the primer sets and our analysis method provided sufficient coverage to identify >1%-frequency sequences throughout the genome. When we evaluated sequences of *pol* genes from 18 treatment-naïve patients' samples, the deep sequencing results were in agreement with Sanger sequencing and identified numerous additional minority mutations. The results suggest that our deep sequencing method would be suitable for identifying within-host viral population dynamics throughout the genome.

Keywords: HIV-1, deep sequencing, drug resistance, error correction, consensus sequence estimation, quasispecies

INTRODUCTION

Knowledge of the genome sequence of human immunodeficiency virus type-1 (HIV-1) is fundamental for improving the clinical outcome of patients infected with HIV-1 and for understanding viral co-evolution within hosts. However, not only between-host HIV-1 genetic diversity and within-host viral heterogeneous population make it difficult to determine the viral

sequences within host. HIV-1 is classified into four groups (M, N, O, and P), and the group that is most widespread globally, M, is further divided into nine subtypes (A, B, C, D, F, G, H, J, and K), with more than 70 circulating recombinant forms (CRFs), according to the Los Alamos HIV Sequence database (http://www.hiv.lanl.gov/), and numerous unique recombinant forms (URFs; Sharp, 2002; Taylor et al., 2008; Hemelaar et al., 2011; Sharp and Hahn, 2011). Genetic diversity between the subtypes ranges from 25 to 35% (Korber et al., 2001), which is extremely high compared to the human population, in which <1% of distinct DNA sequences are distinct (International HapMap, 2003, 2004). This diversity is considered a consequence of HIV-1's short replication period, lack of proofreading machinery, and recombination in viral replication (Robertson et al., 1995; Perelson et al., 1996; Blackard et al., 2002). The genetic diversity of HIV-1 likely influences the effectiveness of antiretroviral therapy (Wainberg and Brenner, 2012) and at least partially prevents the development of curable strategy against HIV-1 infection (Thomson et al., 2002). Moreover, the error-prone replication induces a within-host genetically diverse heterogeneous viral population, recognized as quasispecies (Ojosnegros et al., 2011). The quasispecies are considered a source of drug-resistant or immune escape variants. Within-host minority viruses likely influence clinical outcome (Johnson et al., 2008; Balduin et al., 2009; Geretti et al., 2009; Metzner et al., 2009; Simen et al., 2009; Paredes et al., 2010), although some reports have found no association between treatment failure and minority variants (Peuchant et al., 2008; Jakobsen et al., 2010; Metzner et al., 2010; Stekler et al., 2011).

To improve clinical outcomes and further understand viral co-evolution within-hosts, the HIV-1 RNA genome has been sequenced using the direct Sanger sequencing method. For example, before treatment against HIV-1 infection, the *pol* and *env* V3 regions are sequenced in genotyping resistance tests and tropism tests that predict viral susceptibility to antiretroviral drugs (Smit, 2014). However, analysis of viral polymorphic sequences is limited using Sanger sequencing method. For example, direct Sanger sequencing is inappropriate for analyzing regions containing heterogeneous insertions or deletions, such as *gag* and *env*. Within-host quasispecies population analyses using direct Sanger sequencing can detect low-frequency mutations in only up to 10–20% of the population. In addition, primer design may be troublesome when analyzing sequences of large or multiple segments.

Recently developed next-generation sequencing technologies that output unprecedented amounts of short sequence reads enable the analysis of viral quasispecies in further depth (Willerth et al., 2010; Dudley et al., 2012; Gall et al., 2012; Henn et al., 2012; Gibson et al., 2014; Park et al., 2014). Bench-top deep sequencers Roche/454 GS Junior and Ion PGM, both based on the pyrosequencing method, are applicable for analyses of limited regions of the HIV-1 genome (Dudley et al., 2012; Gibson et al., 2014; Park et al., 2014). Illumina has released another bench-top deep sequencer, MiSeq, based on bridge sequencing technology, which, compared with the aforementioned pyrosequencing platforms, can output large amounts of sequence reads with a lower intrinsic error rate, especially at homopolymeric regions, including the drug-resistance-related reverse transcriptase (RT) K65 codon (Varghese et al., 2010; Loman et al., 2012; Junemann et al., 2013). Here, we have proposed a practical method to analyze viral quasispecies of the HIV-1 near-full-length genome in clinical samples using the Illumina MiSeq deep sequencing method and especially evaluated nucleotide variations in viral sequences of the Pol and the Env V3 encoding regions.

MATERIALS AND METHODS

Plasmid Sample Preparation

To examine analytical biases that may produce misleading results and intrinsic errors in sequence reads from Illumina MiSeq, pNL4-3 ($pNL4\text{-}3_{wt}$) was used as a reference clone. Furthermore, to examine the threshold of deep sequencing in detecting minority mutations in clinical samples, artificially simulated samples were prepared by mixing multiple different clones. For this purpose, three pNL101-based recombinant infectious clones (Neuveut and Jeang, 1996) possessing drug-resistance mutations RT K103N ($pNL101_{rtK103N}$), RT M184V ($pNL101_{rtM184V}$), and integrase (IN) Q148H ($pNL101_{inQ148H}$) were constructed using standard site-directed mutagenesis protocols as described previously (Hachiya et al., 2008; Shimura et al., 2008). Seven ratios of $pNL4\text{-}3_{wt}$: $pNL101_{rtK103N}$: $pNL101_{rtM184V}$: $pNL101_{inQ148H}$ were mixed as follows: (a) 100:0:0:0, (b) 99.97:0.01:0.01:0.01, (c) 99.7:0.1:0.1:0.1, (d) 98.5:0.5:0.5:0.5, (e) 97:1:1:1, (f) 70:10:10:10, and (g) 40:20:20:20.

Clinical Sample Collection and Sanger Sequencing

Fifty-two plasma samples were collected from 33 HIV-1-infected patients who visited the Nagoya Medical Center in Japan from January 2009 to April 2014 (Supplementary Table S1). Forty-five plasma samples collected from 25 patients enrolled in the Japanese Drug Resistance HIV-1 Surveillance Network (Gatanaga et al., 2007; Hattori et al., 2010; Shiino et al., 2014) were also used in this study. Thus, a total of 97 plasma samples obtained from 58 patients were used. To monitor viral quasispecies chronologically, plasma samples were obtained at 10 time points from one patient failing protease inhibitors (PIs) containing regimens. The total RNA was extracted from 200- or 400-μL of the plasma sample using the MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche Diagnostics K.K., Tokyo, Japan). Extracted RNA was eluted in a final volume of 50 μL of elution buffer and used for subsequent analyses. HIV-1 protease (PR) (297 bps; 2253–2549, positions based on HXB2 numbering), RT (720 bps; 2550–3269) and *env* V3 (108 bps; 7110–7217) sequences of each sample were analyzed using the bulk sequencing method as previously reported (Gatanaga et al., 2007; Hattori et al., 2010; Shiino et al., 2014). Drug-resistance mutations in Pol were determined according to the list reported by Shafer et al. (Bennett et al., 2009) and IAS-USA (Wensing et al., 2014). HIV-1 subtypes were determined using phylogenetic analysis with reference sequences recommended by the Los Alamos Database (http://www.hiv.lanl.gov/). Genotypic tropism

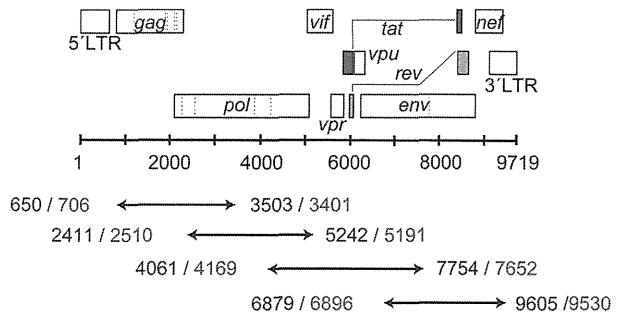


FIGURE 1 | Four overlapping segments covering the HIV-1 near-full-length genome. The blue numbers represent the 5′- and 3′-end positions (HXB2 numbering) for RT-PCR, whereas red numbers show those for nested PCR.

tests were performed using geno2pheno [coreceptor] (http://coreceptor.geno2pheno.org/) with a false-positive rate of 10%.

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

Amplification of the HIV-1 Near-full-length Genome in Clinical Samples

To analyze the full-length HIV-1 genome (excluding LTR regions) using MiSeq, the *gag* to *nef* (8825 bps; 706–9530) region of the genome was amplified in four overlapping segments, as shown in **Figure 1**. The primer sequences used for the amplifications are listed in Supplementary Table S2. RT-PCR was performed using a PrimeScript II High Fidelity One Step RT-PCR Kit (Takara, Shiga, Japan), followed by nested PCR using PrimeSTAR GXL DNA Polymerase (Takara, Shiga, Japan). Finally, four amplified PCR products were combined into one sample with a MultiScreen HTS PCR96 filter plate according to the manufacturer's instructions (Merck Millipore, Billerica, Massachusetts, USA). The purified DNA was eluted into a final volume of 50 μL of distilled water.

Library Preparation and Deep Sequencing with Illumina MiSeq

Viral DNA libraries for deep sequencing were prepared using the Nextera DNA Sample Prep Kit (Illumina K.K.,Tokyo, Japan) according to the manufacturer's instructions. Unamplified DNA of the recombinant clones and amplified DNA obtained from clinical samples were used for the preparation. The prepared library was sequenced using Illumina MiSeq, generating paired-end 2 × 250-bps-long sequence reads. For each run, a maximum of 24 samples were concurrently examined.

Sequence Read Analysis for Deep Sequencing

The sequence reads obtained from Illumina MiSeq were analyzed following three procedures: (I) mapping of the sequence reads onto a reference sequence or a consensus sequence; (II) error correction; and (III) frequency estimation of bases, amino acids, or short (<250-bps-long) fragment sequences. The detailed methods at each step are described below.

(I) Consensus Sequence Estimation and Read Mapping. Sequence read mapping was conducted with the BWA-MEM algorithm implemented in the BWA-0.6.4 program (Li and Durbin, 2009, 2010) using the default setting. We performed four sequence read mapping methods (Supplementary Figure S1A) as follows:

(i) Simple mapping onto an infectious clone sequence. The sequence reads were mapped on a given infectious clone sequence, such as NL4-3, HXB2, JRCSF (GenBank accession no. M38429) in subtype B, and 93JP-NH1(NH1) in CRF01_AE (AB052995).

(ii) Mapping onto consensus sequence derived from *de novo* assembly. Long fragments (>1000 bps) of the HIV-1 near-full-length genome sequence were assembled *de novo* from sequence reads with the VICUNA program (Yang et al., 2012; Malboeuf et al., 2013). The estimated long fragment sequences were mapped onto the HXB2 sequence with the BWA program. Next, the long fragment sequences were connected to construct a near-full-length consensus sequence. Sequence reads were then mapped onto the constructed consensus sequence.

(iii) Mapping onto consensus sequence estimated from iterative mapping (McElroy et al., 2014; Verbist et al., 2015). The first cycle of iterative mapping was performed on a given reference sequence, such as NL4-3, HXB2, JRCSF, or NH1, with BWA using an option of "-B 1" to reduce mismatch penalty for mapping. The majority base at each position was accepted as the consensus base. In cases of insertions or deletions, the longer sequences were always chosen for consensus sequence estimation; i.e., deletions were ignored and insertions were accepted regardless of their prevalence. The consensus sequence estimated in the first cycle was used as a reference sequence of the second cycle of iterative mapping with the BWA program at the default setting. This procedure was repeated nine times to refine a consensus sequence, and, finally, the sequence reads were mapped onto the consensus sequence obtained in the 9th iterative mapping.

(iv) Combination of (ii) and (iii): Sequence reads were mapped onto the consensus sequence estimated from *de novo* assembly using the VICUNA program (Yang et al., 2012; Malboeuf et al., 2013) followed by the iterative mapping (McElroy et al., 2014; Verbist et al., 2015).

(II) Error Correction. We performed error correction using averaged quality score (QS)-values for each reference sequence position. The detailed method is described below.

(III) Frequency Estimation. We estimated the frequency of each base at a given position by counting the number of nucleotides for each base that remained after error correction. We also calculated the occupancy of sequences

within a <250-bps-long range because the maximum length of Illumina MiSeq sequence reads is 250 bp. First, we extracted short fragment sequences from the mapped sequence reads within the targeted range. Then, identical fragment sequences were grouped into a haplotype sequence. The number of the fragment sequences was counted for each haplotype sequence. Next, haplotype sequences including bases of averaged QSs below 20 were removed. The remaining haplotype sequences were used for frequency estimation.

RESULTS

Use of the Consensus Sequence Estimated from *de novo* Assembly or Iterative Mapping Improved the Sequence Read Mapping Results

To examine analytical bias that may produce misleading results and an intrinsic error rate of output sequence reads from Illumina MiSeq, pNL4-3_{wt} was deeply sequenced in quadruplicate. The obtained sequence reads from each of the quadruplicate runs were individually mapped onto the original pNL4-3_{wt} sequence [Schema (i) in Supplementary Figure S1A]. The mapping results demonstrated >5000-fold coverage throughout the HIV-1 genome (**Figure 2**, Supplementary Figure S1B and Supplementary Table S3). To investigate the effect of selected reference sequences on the mapping accuracy and coverage of the sequence reads, two different subtype B sequences, HXB2 (97.4% identical to NL4-3) and JRCSF (93.8%), and one CRF01_AE sequence, NH1 (85.0%), were selected as reference sequences for mapping [Schema (i) in Supplementary Figure S1A]. As shown in **Figure 2**, the mapping coverage by HXB2 (5722- to 23,058-fold) and JRCSF(3038- to 22,570-fold) were comparable to that of NL4-3 (5424- to 21,616-fold), whereas a considerable reduction in coverage was observed when NH1 was used as the reference sequence (0- to 21,064-fold). Thus, selection of the reference sequence is clearly a critical factor for accurate mapping and high coverage of deep sequencing reads. However, for clinical samples, selection of an appropriate reference sequence is problematic because the sample sequences are unknown at the time of deep sequencing. To overcome this problem, we estimated a consensus sequence from *de novo* assembly, iterative mapping, or *de novo* assembly followed by iterative mapping (Yang et al., 2012; Malboeuf et al., 2013; Gibson et al., 2014; McElroy et al., 2014; Verbist et al., 2015) [Schema (ii)–(iv) in Supplementary Figure S1A]. We used the resulting consensus sequence as the reference sequence for mapping, as previously proposed by others (Yang et al., 2012; Malboeuf et al., 2013; Gibson et al., 2014; McElroy et al., 2014; Verbist et al., 2015). The use of the *de novo* assembled consensus sequence provided comparable mapping results (5731- to 23,053-fold) to that of NL4-3. The assembled consensus sequence was analogous to the original NL4-3 reference sequence but included ~10 *nef* mutations, suggesting that *de novo* assembly is likely insufficient to estimate the true consensus sequence. By contrast, iterative mapping or *de novo* assembly followed by iterative mapping estimated a consensus sequence that was identical to the NL4-3 sequence and resulted in the same mapping coverage as that achieved using NL4-3 as the reference sequence. These results suggest that iterative mapping and *de novo* assembly followed by iterative mapping can estimate the true consensus sequence and are most appropriate for sequence read mapping. Hence, in the following sections, we mapped sequence reads from deep sequencing using consensus sequence estimation by *de novo* assembly followed by iterative mapping (Yang et al., 2012; Malboeuf et al., 2013; Gibson et al., 2014; McElroy et al., 2014; Verbist et al., 2015).

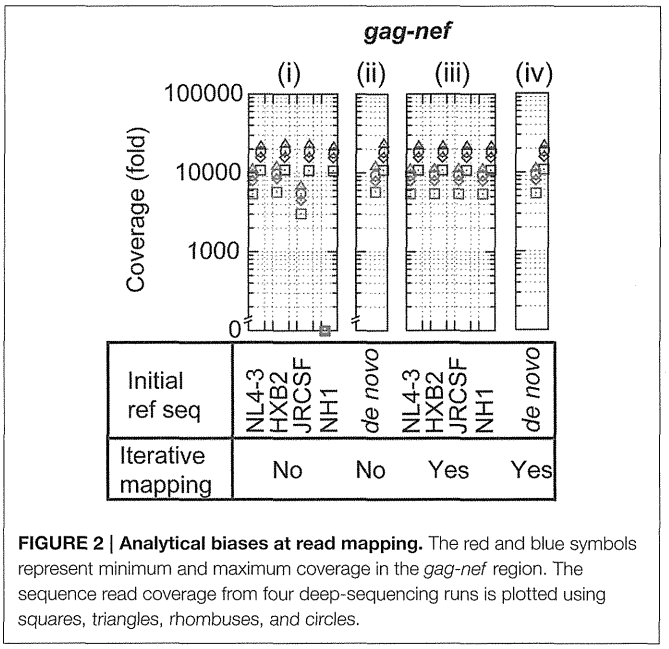


FIGURE 2 | Analytical biases at read mapping. The red and blue symbols represent minimum and maximum coverage in the *gag-nef* region. The sequence read coverage from four deep-sequencing runs is plotted using squares, triangles, rhombuses, and circles.

Our Unique Error-correction Method Reduced the Prevalence of Erroneous Bases Found in Sequence Reads for a Recombinant Clone Below 1% in Each Sequence Position

We also analyzed intrinsic errors in the sequence reads from deep sequencing for pNL4-3_{wt}. As shown in **Figure 3A**, without any error-correction handling, even with clonal pNL4-3_{wt} sequencing results, the erroneous bases occupied a maximum of 6.4% of the population at each reference sequence position. The erroneous bases induced drug-resistance-associated mutations such as IN T66A/I/K and Q148H/K/R at maximums of 1.7 and 1.6% of the population. Considering minority mutation detection by deep sequencing, this error rate is excessive, making minority-mutation determinations inaccurate. In analyzing the patterns of errors, a dominant pattern was substitution (~99.6% in total errors), whereas insertions or deletions were not frequently observed, as previously reported by others (Loman et al., 2012; Junemann et al., 2013). Further examination of substitution patterns revealed that C>A and T>G transversion errors were most frequently observed in the pNL4-3_{wt} sequence

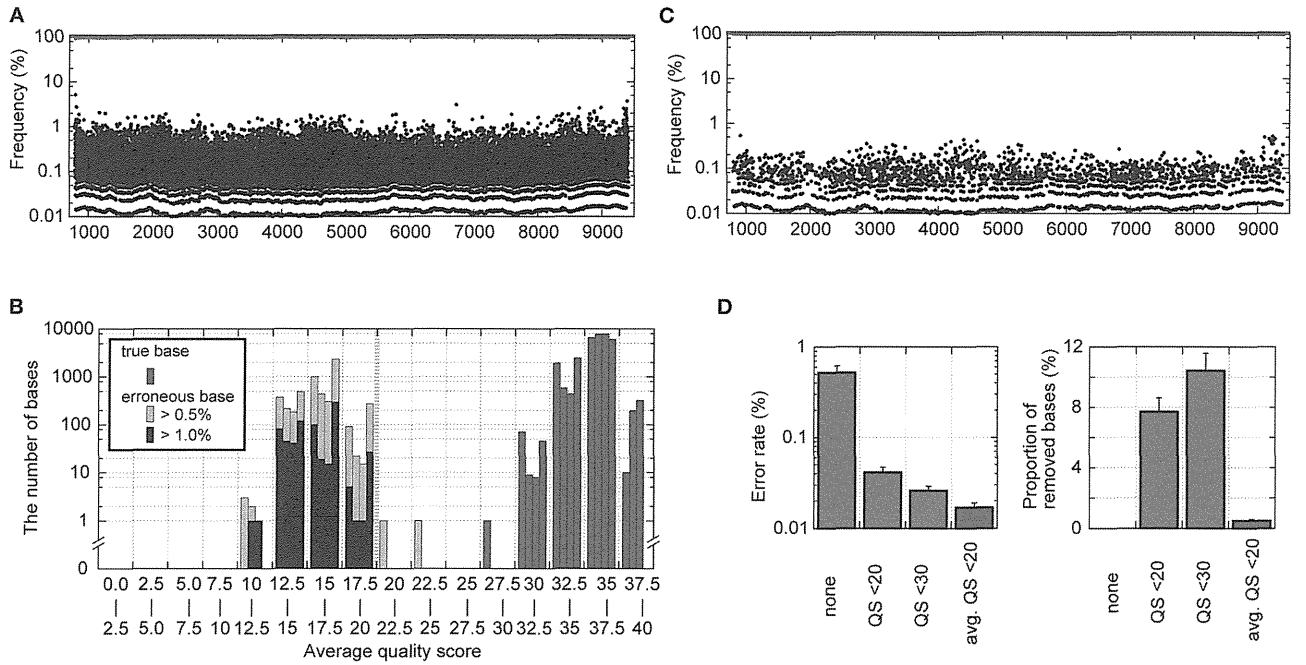


FIGURE 3 | Analytical bias in pNL4-3$_{wt}$ sequence reads from the Illumina MiSeq deep sequencer. (A) Frequencies of nucleotides with true bases (red dots) and with erroneous bases (black dots) throughout the pNL4-3$_{wt}$ genome except for the LTR regions prior to any error correction. **(B)** Histogram of averaged QSs for true bases and erroneous bases at each genome position. The averaged QSs for true bases are shown in red bars, whereas those for erroneous bases occupying >0.5 and >1.0% of the population are represented as gray and black bars, respectively. In each range (x-axis) of the histogram, the four bars representing results from each quadruplicate runs are shown. **(C)** Frequencies of nucleotides with true bases (red dots) and erroneous bases (black dots) throughout the pNL4-3$_{wt}$ genome after error correction removed each base with averaged QSs below 20 at the respective position. **(D)** Error rate after error correction and the proportion of removed nucleotides among all nucleotides using error correction. Results are shown for no error correction, quality-filtering error correction by removing nucleotides with QSs below 20 or 30, and our error correction removing bases of averaged QSs below 20.

reads (Supplementary Figure S2). The predominance of T>G transversion was previously reported for MiSeq (Schirmer et al., 2015) and Illumina Genome Analyzer (Nakamura et al., 2011; Flaherty et al., 2012), both of which are based on the bridge sequencing method, suggesting that the transversion errors might be intrinsic to the apparatuses and the technology.

To improve the accuracy of the sequencing results, we sought to establish a novel error correction method by distinguishing true minority bases from erroneous bases. To differentiate true and erroneous bases, we focused on Phred QS-values of nucleotides in pNL4-3$_{wt}$ sequence reads. QSs of 10, 20, and 30 indicate 90, 99, and 99.9% base-calling accuracy, respectively. Nucleotides with true bases tend to demonstrate high QS-values, whereas low QS-values are associated with erroneous bases (Supplementary Figure S3). When the QSs of nucleotides with true bases were averaged at the respective reference sequence positions, the averaged QS-values for true bases were >25, suggesting >99.7% base-calling accuracies (**Figure 3B**). By contrast, when we focused on bases occupying >1% of the population at the respective positions, the averaged QS-values for erroneous bases were below 20 and were clearly different from those for true bases at a threshold of 20 (**Figure 3B**). Thus, this result indicates that "an averaged QS-value of 20" is a reasonable cut-off threshold to distinguish true and erroneous bases.

Figure 3C shows the results of pNL4-3$_{wt}$ sequencing managed by our novel error-correction method that removed bases with averaged QS-values below 20 at each reference sequence position (Supplementary Figure S3), and the erroneous bases did not occupy >0.54% of the population. Especially, each population of the drug resistant mutations in Pol and erroneous sequences at Env V3 and PR cleavage sites in Gag, Pol, Nef (Shafer and Schapiro, 2008; Fun et al., 2012) was not exceeded 0.2%. The error rates were reduced to 0.017 ± 0.002 from 0.52 ± 0.098% of the raw sequence reads (**Figure 3D**). Of note, our error-correction method removed only <1% of all nucleotides (**Figure 3D**), suggesting selective removal of erroneous bases.

As a comparative method, we applied the simple quality-filtering correction method reported by others (Dudley et al., 2012; Gall et al., 2012; Pessoa et al., 2014) to the same pNL4-3$_{wt}$ sequence reads. The quality-filtering method simply discards any nucleotides with a QS below 20 or 30 as a cut-off value (Supplementary Figure S3). This alternative method was also successful in reducing error rates to 0.041 ± 0.006 and 0.026 ± 0.003% with a cut-off of QS < 20 and QS < 30, respectively (**Figure 3D**). However, the quality-filtering method discarded 7% and 11% of all nucleotides by the QS < 20 and QS < 30 cut-offs, respectively (**Figure 3D**).

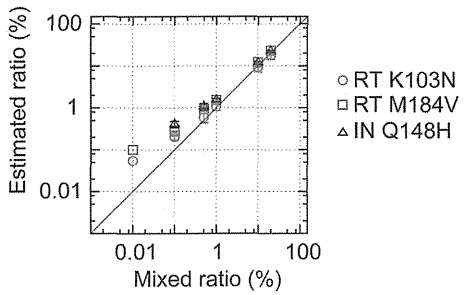


FIGURE 4 | Deep sequencing to detect minority viral populations. pNL4-3_{wt} and pNL101-based mutant clones ($pNL101_{rtK103N}$, $pNL101_{rtM184V}$, and $pNL101_{inQ148H}$) were mixed at seven different ratios: (a) 100:0:0:0, (b) 99.97:0.01:0.01:0.01, (c) 99.7:0.1:0.1:0.1, (d) 98.5:0.5:0.5:0.5, (e) 97:1:1:1, (f) 70:10:10:10, and (g) 40:20:20:20. The estimated frequencies of RT K103N, RT M184V, and IN Q148H are shown in orange circles, red squares, and black triangles, respectively.

Deep-sequencing Analyses Coupled with Our Analysis Method for Distinct Recombinant Clone Mixtures Successfully Detected Minority Mutations with a Prevalence of >1%

To confirm the potential of our mapping and error-correction methods, we performed deep sequencings of mixtures of four recombinant clones, pNL4-3_{wt}, $pNL101_{rtK103N}$, $pNL101_{rtM184V}$, and $pNL101_{inQ148H}$ in triplicate, at seven different ratios: (a) 100:0:0:0, (b) 99.97:0.01:0.01:0.01, (c) 99.7:0.1:0.1:0.1, (d) 98.5:0.5:0.5:0.5, (e) 97:1:1:1, (f) 70:10:10:10, and (g) 40:20:20:20 (**Figure 4**, Supplementary Table S4). We successfully detected three amino acid mutations stably in mixture samples when mutant clones were mixed at ≥0.5% prevalence [(d)–(g)]. For samples (b) and (c), where mutant clones were mixed at 0.01 and 0.1% prevalence, one and two in triplicate tests identified the three mutations, respectively. During the amino acid population analyses, our error-correction removed only ~3% of all three-nucleotide sequences. By contrast, the simple quality-filtering correction method with a QS < 20 or QS < 30 cut-off, removed ~25 or 27% of all three-nucleotide sequences, although the quality filtering correction method also allowed us to stably detect three mutations in samples (d)–(g), where mutant clones were mixed at ≥0.5% prevalence (Supplementary Table S4). Taken together with the aforementioned results on error prevalence, our analysis method enables us to detect amino acid mutations at >1% of the population reproducibly and semi-quantitatively while maximizing usage of the sequence read data.

Hence, in the following sections, errors in sequence reads from deep sequencing were corrected with our error-correction method based on averaged QS-values at a threshold of 20. Furthermore, we used error-corrected >1%-frequency bases, amino acids, or short <250-bps-long fragment sequences.

A Near-full-length Genome Amplification Protocol was Successfully Constructed

We designed primer sets to amplify near-full-length viral RNA genomes in four overlapping segments (**Figure 1**). A total of 97 clinical plasma samples were examined. Phylogeny analyses of *pol* sequences derived from Sanger sequencing indicated that 58 plasma samples from 22 patients and 39 samples from 36 patients contained subtype B and non-subtype B viral RNAs. The results showed that all of the four segments were successfully amplified for the subtype B samples with >200 copies/mL (Supplementary Table S5). By contrast, two subtype B samples with 50 and 65.7 copies/mL were incompletely amplified with missing segments. The same primer sets were tested for the remaining 39 non-subtype B viruses, including 10 subtype C, 10 CRF01_AE, 9 subtype F, and 10 CRF02_AG HIV-1. We successfully amplified all four segments for subtype C, CRF01_AE, subtype F, and CRF02_AG viral genomes from samples with up to 432, 700, 2980, and 1600 copies/mL, respectively (Supplementary Table S5). However, we failed to amplify three subtype F samples below 1240 copies/mL, suggesting lower amplification efficacy for non-subtype B viral genomes than subtype B genomes. In particular, amplification of the *in-env v5* regions in non-subtype B genomes was relatively unsuccessful. This is likely due to more nucleotide mismatches between the primer sequences and non-subtype B viral genome sequences (Supplementary Table S6). Further adjustment of these primers, especially for the *in-env v5* region, is required to improve amplification for non-subtype B viral genomes and to effectively amplify HIV-1 genomes regardless of their subtypes. Consequently, all four segments were successfully amplified from 92 plasma samples.

Subsequently, we deeply sequenced 92 amplified samples. When we analyzed the obtained sequence reads, >95% of the sequence reads were mapped onto the estimated consensus sequence for each sample. Further, as shown in Supplementary Figure S4, >1000-fold coverage of sequence reads were obtained at each position throughout *gag-nef*, except for the 5′-end of *matrix* in six samples, *env* in 1 sample, and the 3′-end of *nef* in 1 sample (Supplementary Table S7). Therefore, each minority nucleotide mutation occupying >1% of the population was confirmed from at least 10 sequence reads. Taken together, the results highlight that amplification with our primer sets followed by deep sequencing enabled us to analyze low-frequency mutations with sufficient sequence read coverage.

Our Deep Sequencing Method Detected more Minority Variants than the Direct Sanger Sequencing Method

To evaluate the potential of detecting quasispecies and minority population using our near-full-length deep sequencing method, we compared the results obtained with the proposed method and the direct Sanger sequencing method. We focused on sequences at PR-RT encoding regions (1017 bps; 2253–3269) from 18 treatment-naïve patients, including 13 subtype B, 4 CRF01_AE, and 1 CRF02_AG viruses (Supplementary Table S1). We achieved both deep and Sanger sequencings from the same amplicons. Only one mismatch was found in one sample (TN07) between the consensus sequences derived using the deep and Sanger sequencing methods (**Figure 5A**), except 216 positions in 18 samples where Sanger sequencing detected

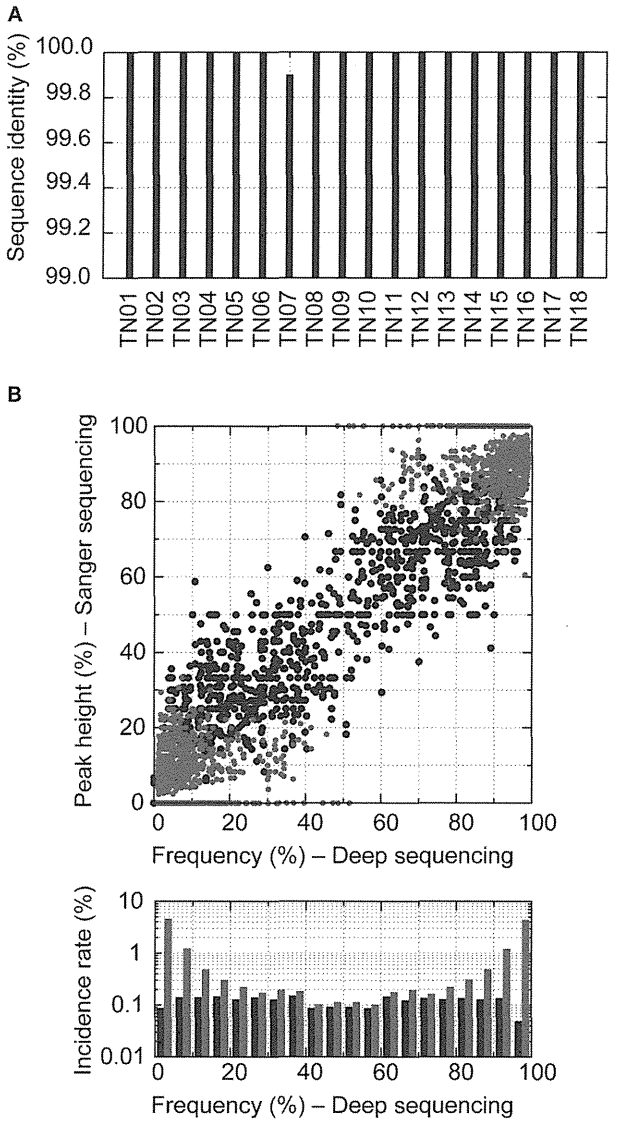


FIGURE 5 | Direct Sanger sequencing and deep sequencing comparison of 18 treatment-naïve samples. (A) Identity of consensus nucleotide sequences at positions where A, T, G, or C was detected by Sanger sequencings between the two methods. **(B)** Correlation of occupancies of bases at the positions where multiple bases at >1% prevalence were detected by deep sequencings between the two methods. For Sanger sequencing, occupancies of bases were estimated from their electropherogram peak heights. Black dots show mixture bases assigned from Sanger sequencing electropherograms. Red dots show mixture bases that were unassigned by Sanger sequencing but were identified by the deep sequencing method. Bottom graphs represent histograms of occupancies detected by the deep sequencing method. The black and red bars show incident rates detected by Sanger sequencings and those detected only by deep sequencing methods, respectively.

mixture bases. Thus, the concordance rate of the two methods was 99.994% [one mismatch in 18,090 (1017 × 18 − 216) positions]. Furthermore, we evaluated the sensitivity of the deep and Sanger sequencing methods in detecting minority populations. For Sanger sequencing, base occupancies were calculated from the electropherogram peak height ratios. As shown in **Figure 5B**, when base occupancies were analyzed at the 216 positions where Sanger sequencing detected mixture bases, a good correlation was observed between the Sanger sequencing method and deep sequencing method (1143 peaks, $R^2 = 0.76$, $P < 0.0001$, single regression analysis; black dots in **Figure 5B**), and all mixture bases detected with Sanger sequencings were identified with the deep sequencing method. However, using the deep sequencing method, we detected an additional 1069 minority bases in all 18 samples that Sanger sequencing failed to recognize (red dots in **Figure 5B**). These results suggest that deep sequencing is more sensitive in detecting minority variants than Sanger sequencing and enables us to analyze quasispecies in clinical samples semi-quantitatively. These additional minority bases contained substitutions conferring drug resistance, and PR M46I (1.1% in TN03, 3.8% in TN04), RT T215S (1.0% in TN01), and RT K219R (1.2% in TN04) were identified. Thus, deep sequencing is likely useful in determining effective treatment regimen in clinical settings.

Our Deep Sequencing Method is Applicable for Detecting Minority X4-tropic Viruses and Examining the Chronological Population Dynamics of Quasispecies

To confirm the clinical advantages of deep sequencing, we applied our deep sequencing method in genotypic tropism testing to determine the co-receptor usage of 18 treatment-naïve patients' samples (Supplementary Figure S5). The deep sequencing method identified heterogeneous V3 sequences in each sample (Supplementary Table S8) and identified nine samples possessing X4-tropic viruses, whereas only 5 samples were identified using Sanger sequencing. As shown in Supplementary Figure S5, all X4 tropic viral sequences in four samples diagnosed using the deep sequencing had less than a 20% minority population (Supplementary Table S8). Thus, the deep sequencing is more sensitive in detecting minority X4-tropic viruses than Sanger sequencings.

To confirm whether our deep sequencing method allows us to dissect quasispecies population dynamics and identify minority drug-resistance mutations relevant to treatment failure, we retrospectively monitored changes in drug-resistance-related mutations in one multi-drug-resistant case (PI-resistant patient #1 in Supplementary Table S1) with M41L/D67N/T69D/M184V/L210W/T215Y in the RT region and M46IL/G73S/I84V/L90M in the PR region (**Figure 6**). We analyzed 10 time points and found that under an EFV-based regimen (time points 3–6), the prevalence of non-nucleoside RT inhibitor (NNRTI)-resistant mutations L100I and K103N increased from <1 to 89.2% and 17.7 to 95.2%, respectively, with viral load relapse from 500 (time point 3) to 5400 copies/mL (time point 5). At time point 6 with 7200 copies/mL, the population of the other NNRTI-resistant mutations, Y181C (33.3%) and G190S (36.4%) increased, whereas the prevalence of L100I and K103N decreased (58.3 and 59.5%). These NNRTI-mutations became undetectable after the regimen was switched

to LPV/r-based therapy (time points 7–10). Emergence of PR I54L followed by I54V was also correlated with relapse under LPV/r-based therapy. Thus, the dynamic population movements of drug-resistance mutations were successfully monitored in detail using our deep sequencing method. In addition to drug-resistance mutations, we also found a population possessing the NC/p1 cleavage-site mutation AP2V was fluctuating with the regimen switches, when we analyzed the sequences at all 11 PR cleavage sites in Gag, Pol, and Nef (Shafer and Schapiro, 2008; Fun et al., 2012). The mutation was first identified as the major population at time point 1, when SQV-based therapy was in progress. Subsequently, the mutation became a minority with EFV-based therapy, and revived as the majority with LPV/r-based therapy.

Although clinical significance was not confirmed in these two analyses of tropism testing and drug resistance mutation monitoring, our results suggest that our deep sequencing method for clinical sample analysis generates more data for understanding within-host viral co-evolutions such as tropism drifting and selection of antiretroviral resistances.

DISCUSSION

In this study, we have proposed a practical method to analyze viral quasispecies of the HIV-1 near-full-length genome in clinical samples using the Illumina MiSeq deep sequencing method (Supplementary Figure S6). Sequence data with low error rates are crucial for accurately analyzing minority populations and genetic diversity of HIV-1. We applied a unique error correction to minimize the effect of artificial errors and facilitate HIV-1 genome analysis using the Illumina bridge sequencing technology. Of note, Illumina bridge sequencing produces $0.52 \pm 0.098\%$ reading errors, which is significantly greater than sequencing platforms using high-fidelity polymerases (Cline et al., 1996; Palmer et al., 2005), but lower than pyrosequencing methods exhibiting high error rates at homopolymeric regions (Varghese et al., 2010; Dudley et al., 2012; Loman et al., 2012; Di Giallonardo et al., 2013; Junemann et al., 2013; Gibson et al., 2014). Our sequencing analysis of the infectious clone pNL4-3_{wt} indicated that the averaged QS-value is a reasonable guide to distinguish true and erroneous bases. One advantage of our error-correction method based on the averaged QS-value is that it removes significantly fewer nucleotides than quality-filtering error correction methods previously reported (Dudley et al., 2012; Gall et al., 2012; Pessoa et al., 2014), which increases the opportunity for detecting minority mutations.

In addition, sequencing analysis of the pNL4-3_{wt} clone suggested that reference sequence choice is critical for accurate and efficient sequence read mapping. To select an appropriate reference sequence in clinical sample analyses, we found that consensus sequence estimated from sequence reads is applicable as the reference sequence for the mapping as reported previously by others (Yang et al., 2012; Malboeuf et al., 2013; McElroy et al., 2014; Verbist et al., 2015), and that *de novo* assembly followed by iterative mapping [Schema (iv) in Supplementary Figure S1A] precisely estimates consensus sequence. During sequence analysis at PR-RT encoding regions of treatment-naïve patients' samples, *de novo* assembly followed by iterative mapping estimated the same consensus sequence as Sanger sequencing, except for one mutation in TN07 (**Figure 5**). By contrast, another estimation method, *de novo* assembly [Schema (ii) in Supplementary Figure S1A] inferred consensus sequences with two mutations for TN11 and another mutation for TN07.

When we performed phylogeny analyses of these estimated near-full-length consensus sequences for each clinical sample, their phylogenetic tree showed concordant subtypes to those based on their *pol* sequences by Sanger method (Supplementary Figure S7), except two samples (Non-subtype B samples #23 and #26 in Supplementary Table S1). Although these two samples were classified into CRF02_AG from the *pol* and of subtype A or G from the near-full-length sequences, this is likely due to analyzed sequence lengths and/or recombination breakpoint positions within CRF02_AG. Furthermore, phylogenetic clusters were found among samples from each drug-resistant patient. Samples from a partner pair (TN12 and TN13) diagnosed in our hospital were also phylogenetically close to each other. Taken together with high sequence identities of the Pol and Env V3 encoding regions between Sanger and our methods, these results suggest that our method may estimate consensus sequences throughout near-full-length regions accurately.

With our error-correction and mapping methods, we can obtain benefit of large genome information from the bridge sequencing. Our method enables both in-depth and semi-quantitative quasispecies analyses (**Figure 5** and Supplementary Figure S5). Although we especially evaluated sequences at the Pol and Env V3 encoding regions in this study, our method would be applicable for quasispecies analyses at the other regions such at PR cleavage sites as shown in **Figure 6**. This is an advantage of our method that is applicable to analyze sequences throughout near-full-length genomes in depth at a run, unlike Sanger or allele specific sequencing methods. Of note, we successfully amplified genomes from low viral load samples using our designed primer sets. Therefore, when patient's viral load increases above the detection limit, our method might be helpful for early detection of drug resistant mutations. Collecting and analyzing genome date using our methods will lead to a comprehensive understanding of unknown mechanisms of resistance acquisition and treatment failure, such as the recent finding demonstrating the importance of the Env cytoplasmic tail mutation in PI resistance (Rabi et al., 2013). Moreover, our method would also be applicable to examine whether drug resistant variants are persistent as proviral DNA, although further assessment is required. Combination of *in vitro* resistance induction experiments or *in vivo* infection of HIV-1 relatives to animal models with our method would help recognize drug resistant machinery or viral evolution. However, there are at least two limitations with our analysis method. The first is due to short sequence reads. Because sequence reads in our study were up to 250-bps long, it was difficult to evaluate interferences of two or more mutations that locate more than 250-bps distant positions. Despite the limitation, our deep sequencing method could help

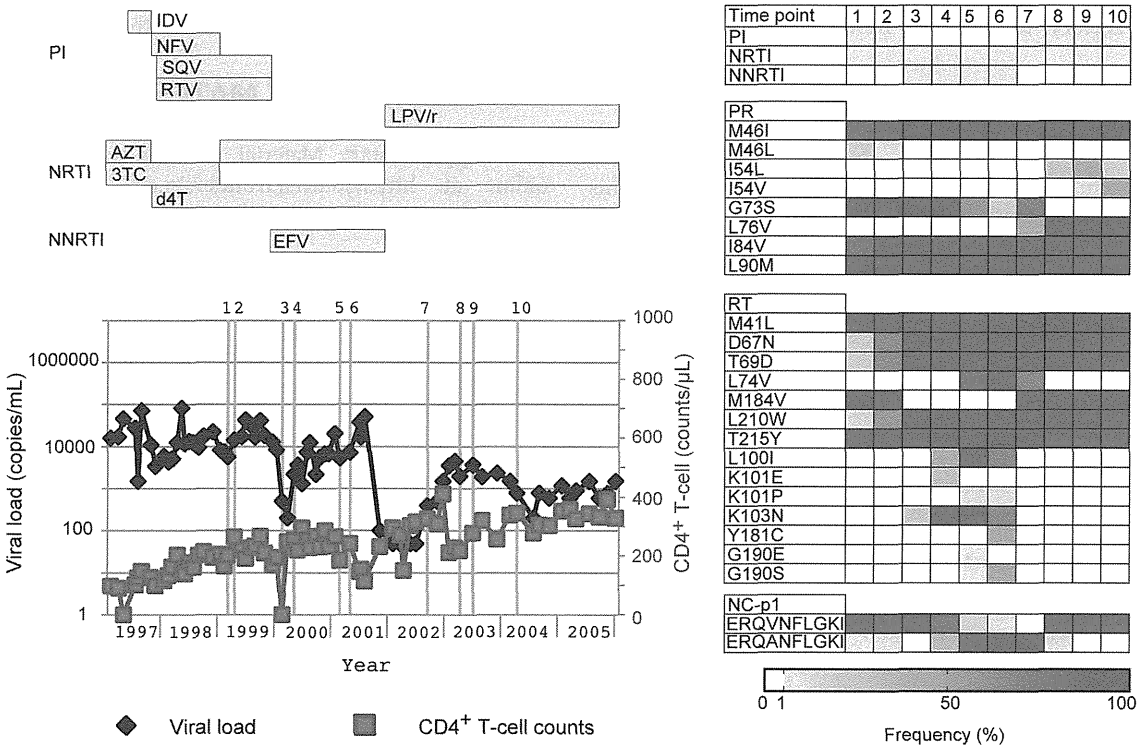


FIGURE 6 | Application of our deep sequencing method in examination of chronological population changes of drug-resistance mutations. Chronological changes in viral loads, $CD4^+$ T-cell counts, and occupancies of drug-resistance mutations for a patient who failed a PI-containing regime. In the left graph, viral loads and $CD4^+$ T-cell counts are shown in black and light red lines, respectively. Collection dates of analyzed samples are highlighted in orange perpendicular lines. Over the orange line, the identification numbers for the collection dates are also shown. At the top of the graph, periods when drugs were used are shown in horizontal bars. Furthermore, population dynamics of resistance related mutations are shown on the right of the graph. The occupancies of the mutations are highlighted with colors according to the bottom bar. PI, PR inhibitor; NRTI, nucleoside/nucleotide RT inhibitor; NNRTI, non-nucleoside RT inhibitor. IDV, indinavir; NFV, nelfinavir; SQV, saquinavir; RTV, ritonavir; LPV/r, lopinavir boosted with ritonavir; AZT, azidothymidine; 3TC, lamivudine; d4T, stavudine; EFV, efavirenz.

obtain hints to know co-evolution within the genome, like mutations in PR and its cleavage sites, in combination with clonal sequencings or haplotype inference by recently proposed some bioinformatics algorithms (Beerenwinkel and Zagordi, 2011; Beerenwinkel et al., 2012; Prosperi et al., 2013; Giallonardo et al., 2014; Schirmer et al., 2014; Jayasundara et al., 2015). The second limitation is attributable to limited stocks of plasma viral RNAs. In several clinical samples, cDNA was amplified from 2.5 μL of extracted viral RNA for each of the four segments, which theoretically contained less than 100 copies of viral RNA, as the original 200 μL plasma had a viral load of <5000 copies/mL. This limitation alerts that the results might be less heterogeneous population than in reality, although we attempted to reduce this risk by triplicate genome amplifications for deep sequencing of each sample. To address this second limitation, in addition to triplicate genome amplification, we must consider increasing several parameters, including the amount of templates, viral RNA, plasma, PCR volume, interestingly retrogressing direction of downsizing sequence technology progress in the last decade.

In conclusion, we devised a data management method and library preparation protocol to analyze quasispecies throughout the HIV-1 near-full-length genome using Illumina MiSeq bentchtop deep sequencing technology. Using deep-sequencing technology with larger genome datasets to precisely analyze minority drug-resistance mutations may improve the efficacy of antiretroviral therapy management in clinical settings.

ACCESSION NUMBER

The data sets analyzed in this study have been deposited in the DNA Data Bank of Japan (DDBJ) under Bioproject accession number PRJDB3502.

AUTHOR CONTRIBUTION

Conceived and designed the experiments: HO, MM, KM, WS. Performed the experiments: HO, MM, KM, AH, JH, YK, YY, YI, WS. Analyzed the data: HO, MM, KM, WS. Contributed reagents/materials/analysis tools: YY, YI, WS. Wrote the paper: HO, MM, KM, WS.

FUNDING

This study was supported by a Grant-in-Aid for AIDS research from the Ministry of Health, Labour, and Welfare of Japan

(H25-AIDS-004) and the Research Program on HIV/AIDS from the Japan Agency for Medical Research and Development, AMED.

ACKNOWLEDGMENTS

We are grateful to all the patients who participated in this study. We thank Ms. Reiko Okazaki, Urara Shigemi and Masumi Hosaka for sample preparation. pNL101 was kindly provided by K.-T. Jeang (National Institutes of Health, Bethesda, MD).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.01258

REFERENCES

Balduin, M., Oette, M., Daumer, M. P., Hoffmann, D., Pfister, H. J., and Kaiser, R. (2009). Prevalence of minor variants of HIV strains at reverse transcriptase position 103 in therapy-naive patients and their impact on the virological failure. *J. Clin. Virol.* 45, 34–38. doi: 10.1016/j.jcv.2009.03.002

Beerenwinkel, N., Gunthard, H. F., Roth, V., and Metzner, K. J. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front. Microbiol.* 3:329. doi: 10.3389/fmicb.2012.00329

Beerenwinkel, N., and Zagordi, O. (2011). Ultra-deep sequencing for the analysis of viral populations. *Curr. Opin. Virol.* 1, 413–418. doi: 10.1016/j.coviro.2011.07.008

Bennett, D. E., Camacho, R. J., Otelea, D., Kuritzkes, D. R., Fleury, H., Kiuchi, M., et al. (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS ONE* 4:e4724. doi: 10.1371/journal.pone.0004724

Blackard, J. T., Cohen, D. E., and Mayer, K. H. (2002). Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clin. Infect. Dis.* 34, 1108–1114. doi: 10.1086/339547

Cline, J., Braman, J. C., and Hogrefe, H. H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* 24, 3546–3551. doi: 10.1093/nar/24.18.3546

Di Giallonardo, F., Zagordi, O., Duport, Y., Leemann, C., Joos, B., Kunzli-Gontarczyk, M., et al. (2013). Next-generation sequencing of HIV-1 RNA genomes: determination of error rates and minimizing artificial recombination. *PLoS ONE* 8:e74249. doi: 10.1371/journal.pone.0074249

Dudley, D. M., Chin, E. N., Bimber, B. N., Sanabani, S. S., Tarosso, L. F., Costa, P. R., et al. (2012). Low-cost ultra-wide genotyping using Roche/454 pyrosequencing for surveillance of HIV drug resistance. *PLoS ONE* 7:e36494. doi: 10.1371/journal.pone.0036494

Flaherty, P., Natsoulis, G., Muralidharan, O., Winters, M., Buenrostro, J., Bell, J., et al. (2012). Ultrasensitive detection of rare mutations using next-generation targeted resequencing. *Nucleic Acids Res.* 40:e2. doi: 10.1093/nar/gkr861

Fun, A., Wensing, A. M., Verheyen, J., and Nijhuis, M. (2012). Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology* 9:63. doi: 10.1186/1742-4690-9-63

Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., et al. (2012). Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* 50, 3838–3844. doi: 10.1128/JCM.01516-12

Gatanaga, H., Ibe, S., Matsuda, M., Yoshida, S., Asagi, T., Kondo, M., et al. (2007). Drug-resistant HIV-1 prevalence in patients newly diagnosed with HIV/AIDS in Japan. *Antiviral Res.* 75, 75–82. doi: 10.1016/j.antiviral.2006.11.012

Geretti, A. M., Fox, Z. V., Booth, C. L., Smith, C. J., Phillips, A. N., Johnson, M., et al. (2009). Low-frequency K103N strengthens the impact of transmitted drug resistance on virologic responses to first-line efavirenz or nevirapine-based highly active antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* 52, 569–573. doi: 10.1097/QAI.0b013e3181ba11e8

Giallonardo, F. D., Topfer, A., Rey, M., Prabhakaran, S., Duport, Y., Leemann, C., et al. (2014). Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res.* 42:e115. doi: 10.1093/nar/gku537

Gibson, R. M., Meyer, A. M., Winner, D., Archer, J., Feyertag, F., Ruiz-Mateos, E., et al. (2014). Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. *Antimicrob. Agents Chemother.* 58, 2167–2185. doi: 10.1128/AAC.02710-13

Hachiya, A., Kodama, E. N., Sarafianos, S. G., Schuckmann, M. M., Sakagami, Y., Matsuoka, M., et al. (2008). Amino acid mutation N348I in the connection subdomain of human immunodeficiency virus type 1 reverse transcriptase confers multiclass resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. *J. Virol.* 82, 3261–3270. doi: 10.1128/JVI.01154-07

Hattori, J., Shiino, T., Gatanaga, H., Yoshida, S., Watanabe, D., Minami, R., et al. (2010). Trends in transmitted drug-resistant HIV-1 and demographic characteristics of newly diagnosed patients: nationwide surveillance from 2003 to 2008 in Japan. *Antiviral Res.* 88, 72–79. doi: 10.1016/j.antiviral.2010.07.008

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 25, 679–689. doi: 10.1097/QAD.0b013e328342ff93

Henn, M. R., Boutwell, C. L., Charlebois, P., Lennon, N. J., Power, K. A., Macalalad, A. R., et al. (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* 8:e1002529. doi: 10.1371/journal.ppat.1002529

International HapMap, C. (2003). The international hapmap project. *Nature* 426, 789–796. doi: 10.1038/nature02168

International HapMap, C. (2004). Integrating ethics and science in the international hapmap project. *Nat. Rev. Genet.* 5, 467–475. doi: 10.1038/nrg1351

Jakobsen, M. R., Tolstrup, M., Sogaard, O. S., Jorgensen, L. B., Gorry, P. R., Laursen, A., et al. (2010). Transmission of HIV-1 drug-resistant variants: prevalence and effect on treatment outcome. *Clin. Infect. Dis.* 50, 566–573. doi: 10.1086/650001

Jayasundara, D., Saeed, I., Maheswararajah, S., Chang, B. C., Tang, S. L., and Halgamuge, S. K. (2015). ViQuaS: an improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing. *Bioinformatics* 31, 886–896. doi: 10.1093/bioinformatics/btu754

Johnson, J. A., Li, J. F., Wei, X., Lipscomb, J., Irlbeck, D., Craig, C., et al. (2008). Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naive populations and associate with reduced treatment efficacy. *PLoS Med.* 5:e158. doi: 10.1371/journal.pmed.0050158

Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* 58, 19–42. doi: 10.1093/bmb/58.1.19

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30, 434–439. doi: 10.1038/nbt.2198

Malboeuf, C. M., Yang, X., Charlebois, P., Qu, J., Berlin, A. M., Casali, M., et al. (2013). Complete viral RNA genome sequencing of ultra-low copy