Supplemental Material

Nanopore sequencing for characterization of HIV-1 recombinant forms

Mikiko Mori[a,b], Hirotaka Ode[a], Mai Kubota[a], Yoshihiro Nakata[a,b], Takaaki Kasahara[a,b], Urara Shigemi[a], Reiko Okazaki[a], Masakazu Matsuda[a], Kazuhiro Matsuoka[a], Atsuko Sugimoto[a], Atsuko Hachiya[a], Mayumi Imahashi[a], Yoshiyuki Yokomaku[a], and Yasumasa Iwatani[a,b]#

[a]Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan

[b]Division of Basic Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan

TABLE S1. List of primer sets for genome amplification by RT-PCR (1st) and nested PCR (2nd).

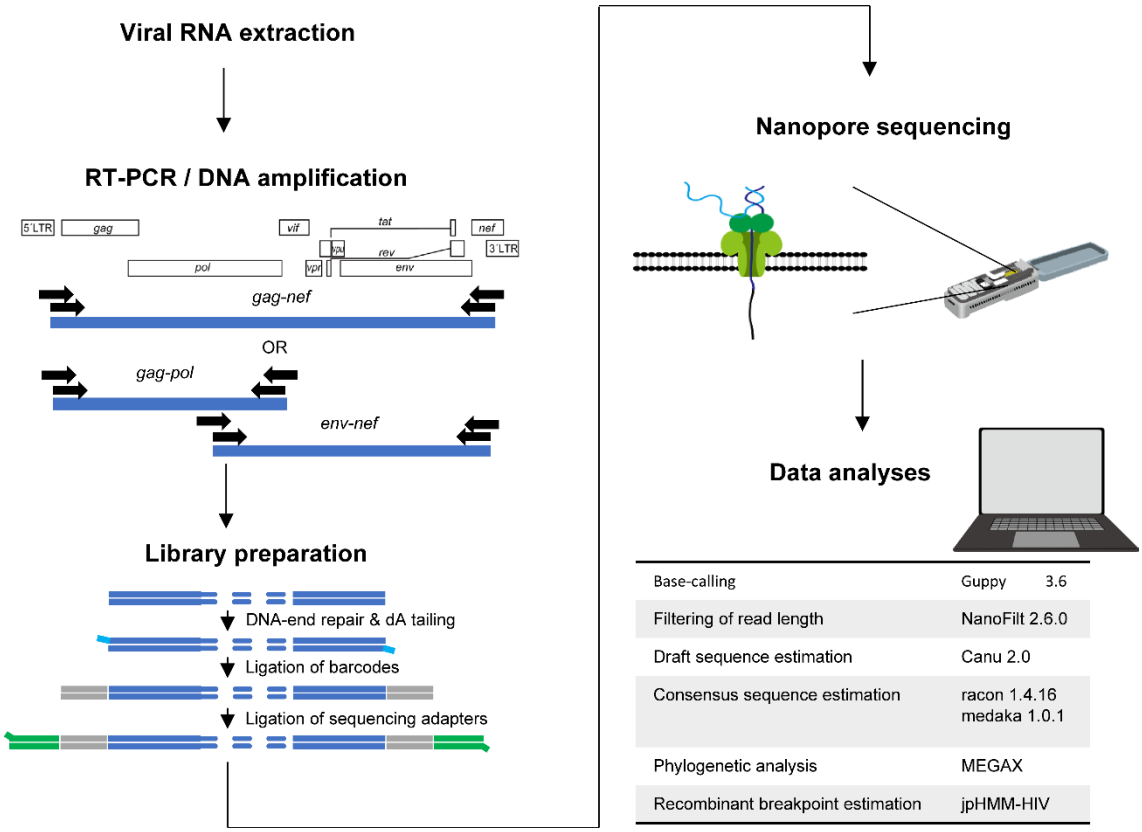
			sequences (5′ to 3′)	length (base)	Position number of nucleotide (5′ to 3′)
gag-nef	1st	F *	ATCTCTAGCAGTGGCGCCCGAACAG	25	625-649
	2nd	F	CTCTCTCGACGCAGGACTCGGCTTG	25	681-705
	1st	R	CACTCAAGGCAAGCTTTATTGAGGC	25	9,630-9,606
	2nd	R	GGTCTAACCAGAGAGACCCAGTACAG	26	9,556-9,531
gag-in	1st	F	ATCTCTAGCAGTGGCGCCCGAACAG	25	625-649
	2nd	F	CTCTCTCGACGCAGGACTCGGCTTG	25	681-705
	1st	R	CCTGTATGCAGACCCCAATATG	22	5,264-5,243
	2nd	R	CCTAGTGGGATGTGTACTTCTGAACTTA	28	5,219-5,192
in-nef	1st	F	CAGACTCACAATATGCATTAGG	22	4,039-4,060
		F †	CAGACTCACAGTATGCATTAGG	22	4,039-4,060
	2nd	F	CTGGCATGGGTACCAGCACACAA	23	4,146-4,168
	1st	R	CACTCAAGGCAAGCTTTATTGAGGC	25	9,630-9,606
	2nd	R	GGTCTAACCAGAGAGACCCAGTACAG	26	9,556-9,531

* F and R denote forward and reverse primers, respectively.

†The primer was newly designed in this study.

The numbers represent the nucleotide positions in the HXB2 reference sequence.

Supplementary Figures



Base-calling	Guppy 3.6
Filtering of read length	NanoFilt 2.6.0
Draft sequence estimation	Canu 2.0
Consensus sequence estimation	racon 1.4.16 medaka 1.0.1
Phylogenetic analysis	MEGAX
Recombinant breakpoint estimation	jpHMM-HIV

FIG S1 A flowchart of the nanopore sequencing protocol applied to the near-full-length HIV-1 genome in this study.

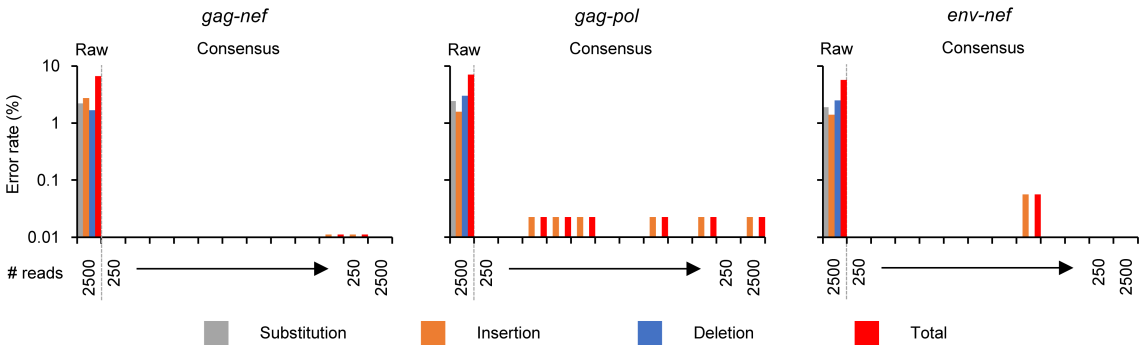


FIG S2 Error rate (%) among raw reads and consensus sequences generated by nanopore sequencing. The proviral DNA, pNL4-3, was used for analysis. Individual error patterns (substitutions, insertions, deletions, and their total) are colored. Consensus sequences were estimated repeatedly from 10 different sets of 250 reads or from all reads (2,500 reads). No bar represents no error.

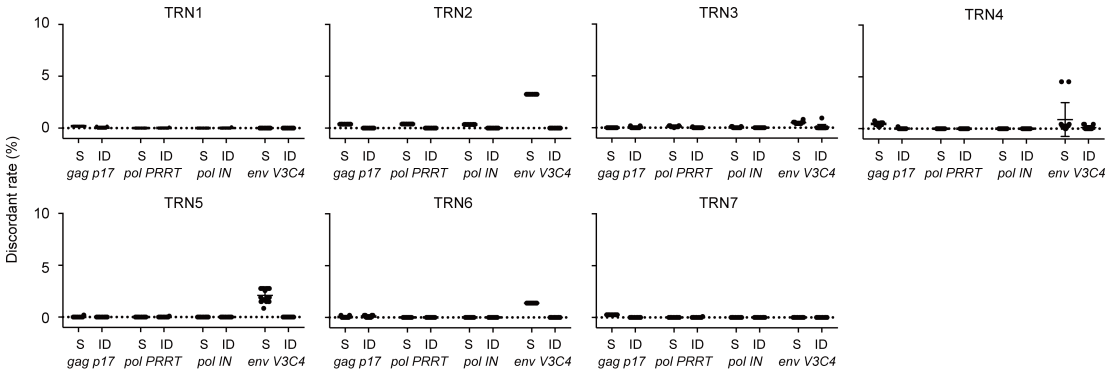


FIG S3 Ratio of mismatched bases between Sanger and nanopore sequencing. The viral genome sequences in the *gag p17* (positions 790 to 1,185 relative to the reference HXB2), *pol PRRT* (2,253 to 3,269), *pol IN* (4,230 to 5,093), and *env V3C4* (7,114 to 7,589) regions were compared. The ratios at the nucleotide positions are plotted with dots. Horizontal lines indicate the means with SDs shown as error bars. S, substitution. ID, insertion or deletion.

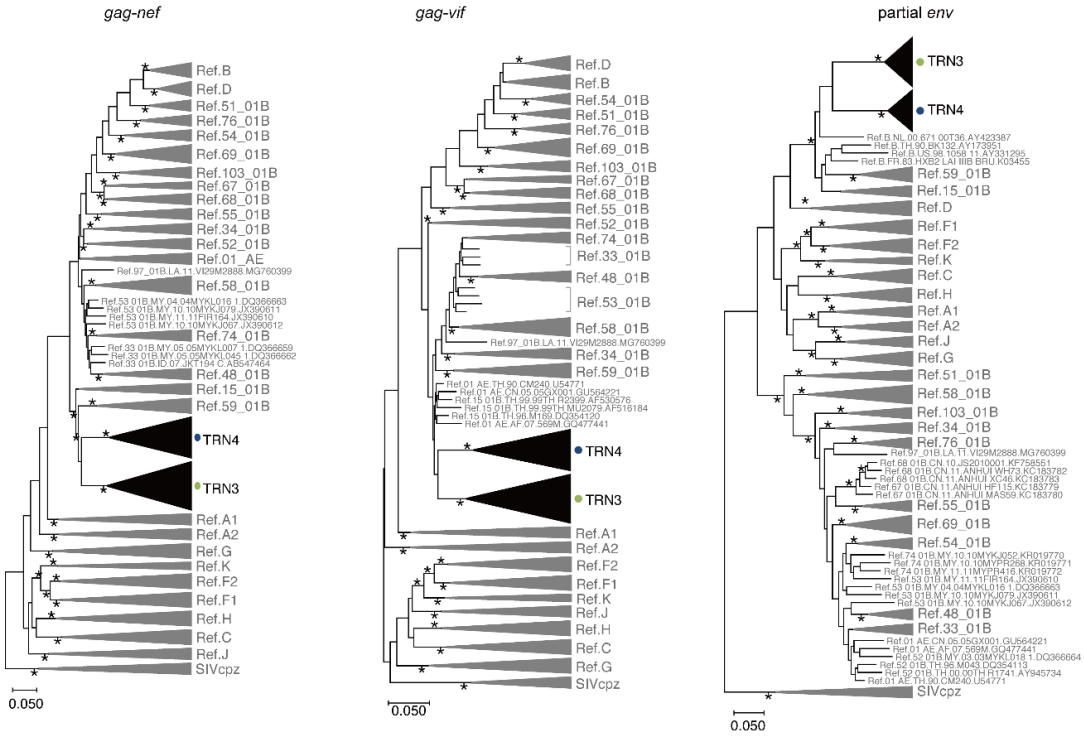


FIG S4 Phylogenetic relationships of the recombinant forms between subtype B and CRF01_AE. Pangenomic sequences and coding sequences of the *gag-vif* (positions 790 to 5,619 nt) and *env* (positions 6,300 to 8,200 nt) regions were analyzed. A maximum likelihood tree was constructed. Three SIVcpz sequences (GenBank IDs: DQ373064, DQ373063, and EF535994) are used as outliers. Branches with bootstrap values (based on 500 replicates) of at least 0.95 are highlighted with asterisks.

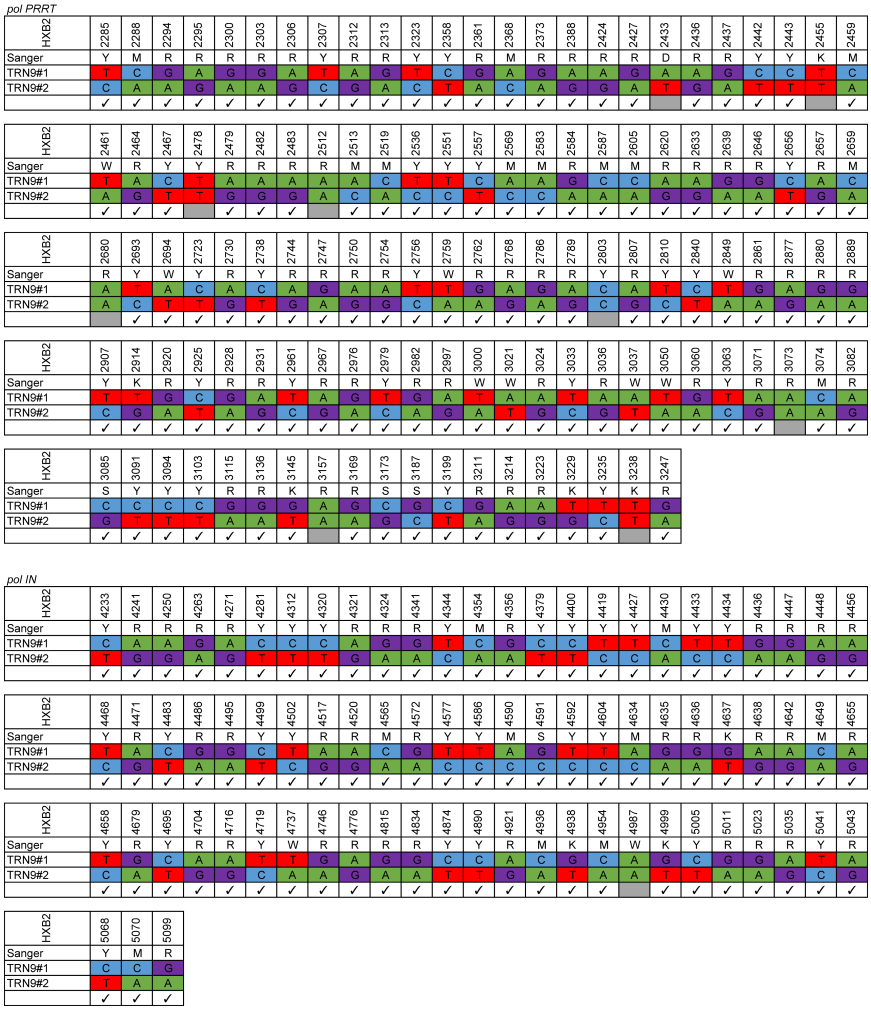


FIG S5 Detection of haplotype sequences in the *pol* region in TRN9 by nanopore sequencing. Two or more mixed bases at the 119th position in the *pol PRRT* gene (1,017 nt) and at the 78th position in the *pol IN* gene (864 nt) were detected. Subsequent analysis by nanopore sequencing revealed two haplotype sequences in the regions. Major base contents of the positions (92% and 99% at positions 110 nt and 77 nt, respectively) were separately assigned to the two haplotype sequences (marked by "✓"). Minor base contents (gray background) were situated where the bases were consistently assigned between the two haplotype sequences through both nanopore sequencing and Sanger sequencing. D = A, G, or T.