

主論文の要旨

Nanopore Sequencing for Characterization of HIV-1 Recombinant Forms

〔 Nanopore sequencingを利用したHIV-1のゲノム組換え体の解析 〕

名古屋大学大学院医学系研究科 総合医学専攻
免疫不全統御学講座 免疫不全統御学分野

(指導：岩谷 靖雅 教授)

森 美喜子

【緒言】

後天性免疫不全症候群(エイズ)の原因である HIV-1 は、ウイルス粒子に約 10,000 塩基のウイルスゲノム RNA を 2 本持つウイルスである。HIV-1 の大きな特徴の 1 つとして、複製エラーおよび組換えによるゲノム変異が高頻度に生じることが挙げられる。そのため、これまでに感染者由来の多様な配列が観察されている。HIV-1 ゲノム配列の遺伝系統学的な分類では、主流のグループ M に含まれる 10 サブタイプとそれらサブタイプ間の組換え体(Recombinant Form : RF)が報告されている。組換え体は流行型(Circulating RF : CRF)と孤発型(Unique RF : URF)に分類され、2022 年 9 月時点では 132 種の HIV-1 CRF が同定されている。近年、特に CRF の報告が増加しており、今後さらに RF が検出されると予想される。しかし、RF を検出するには、重複感染によるウイルス混合と区別するためにクローニングや限界希釈法などの事前処理が必要となるため、解析は容易ではない。そこで本研究では、長鎖 1 分子配列の解析が可能な Nanopore sequencing を活用した、簡便な HIV-1 ゲノム RF の解析系の構築を行った。

【対象および方法】

2016 年から 2020 年における独立行政法人国立病院機構名古屋医療センターの 373 症例について、Sanger 法による HIV-1 薬剤耐性検査の結果から HIV-1 RF あるいは重複感染が疑われた 9 症例を対象とした。薬剤耐性検査ではウイルスゲノム上の 4 領域を解析しており、それら 4 領域が不一致のサブタイプを示した場合に RF を疑った。血漿からウイルス RNA を抽出し、全長あるいは 2 断片として近全長ゲノム領域を RT-PCR により増幅した。MinION(Oxford Nanopore 社)を使用して配列情報を回収し、各検体由来ウイルスのコンセンサス配列を決定した(Fig. 1)。薬剤耐性検査で得られた Sanger 法による配列情報を比較対象として利用した。

【結果】

Nanopore sequencing により 2,500 以上のリード配列を取得し、期待値に近い長さのリード配列を使用して長鎖ゲノム解析を行った。臨床サンプルを解析する前に、分子クローン HIV-1(NL4-3 株)を用いて Nanopore sequencing のエラー(塩基の置換、挿入および欠失)の出現頻度を算出した。250 リードを 1 セットとして計 10 セットの配列解析を行ったところ、各セットで得られたコンセンサス配列のエラー率は約 1/5,000 となり、この手法により高精度のコンセンサス配列情報を取得できることが確認できた(Fig. 2)。次に、RF もしくは重複感染が疑われた 9 症例のコンセンサス配列を同手法で決定したところ、7 つが HIV-1 RF、2 つが重複感染と判定された。薬剤耐性検査の結果から RF と疑われた症例において、薬剤耐性検査で解析した各 4 領域のサブタイプは Nanopore sequencing で得られたコンセンサス配列のサブタイプと一致していた(Fig. 3)。各症例の近全長ゲノムにおけるサブタイプのモザイク構造を Fig. 4 に表す。また、コンセンサス配列はゲノムの中で多様性が最も大きい *env* 可変領域を除き、Sanger 法の配列と高い一致率を示した(Fig. 5)。一方、異種サブタイプウイルスの重複

感染疑いの2症例では、それぞれのサブタイプの混合配列として分離・取得することができ、TRN9に関してはRF同士の重複感染と判定された(Fig. 6と7)。

【考察】

Nanopore sequencing から得られた HIV-1 配列データを利用して新しい解析系を構築した。

本解析法のポイントとしては、① 標的領域の end-to-end のリード配列の抽出により、異なるリード配列数で生じるバイアスやアライメントの誤差を低減しコンセンサス配列を推定したこと、② 異なる 250 本のリード配列を用いてコンセンサス配列の推定を各 10 回行うことで再現性と重複感染事例の場合のマイナーバリエーションの検出感度が向上したこと、が挙げられる。その結果、本解析法の誤差を 0.011~0.056% に低減させることができた。また、限界希釈法などで1分子のゲノムを分離することなく、臨床サンプルから抽出したウイルス RNA をそのまま利用して増幅し、増幅したアンプリコンを Nanopore sequencing により解析した。その際、HIV-1 RF と重複感染例の混合ウイルスを in-silico で分離した点が本解析法のユニークな点である。

本研究で使用したナノポアシーケンサー MinION は重量 87g の小型装置である。MinION に装着したフローセルにサンプルを数滴滴下し、高速 USB ケーブルを介してパソコンに接続することでゲノム解析をスムーズに行うことが可能であった。フローセルには1分子の DNA だけが通過できるポアという構造が無数に存在しており、DNA がポアを通過した際に発生した電気的シグナルを検出してヌクレオチドの順序を判定するというメカニズムである。Nanopore sequencing は開発途上国や北極圏、国際宇宙ステーションでも実用されており、作業環境が実験室に限定されない。近年では新しい RF が次々と出現し HIV-1 の遺伝的多様化が進んでいるため、HIV-1 の全長ゲノムシーケンスは疫学的伝播の追跡や新規 RF の同定に推奨されると考えられる。既報の研究では、ある種の薬剤耐性関連変異が HIV-1 薬剤耐性検査の標的配列領域外に存在することが報告されている。例えばウイルスプロテアーゼが認識する *Gag* や *Gag-Pol* の切断部位の変異、*env* 領域や *polypurine tract* のインテグラーゼ阻害剤の耐性に関連する変異などが挙げられる。そのため、長鎖1分子配列の解析が有用であると考えられている。ただし、本研究には2つの留意点がある。まず、現時点において Nanopore sequencing の技術は、*homopolymer* 領域や *short tandem repeat* 領域に塩基挿入のエラーがまれに発生する弱点がある。HIV-1 は *pol* や *polypurine tract* 領域にホモポリマー配列を有することが多いため、1分子のゲノム解析において意図しない塩基挿入や欠失がないか確認するための検証作業が必要となる。第二に、Nanopore sequencing で得られるマイナー配列がコンセンサス配列構築時にメジャー配列によって覆い隠されることがある。そのため、結果的に得られた配列多様性(特にマイナー HIV-1 変異体に関して)は本来の配列を反映していない可能性がある。これらの留意点をさらに改善することが今後の課題である。

【結語】

Nanopore sequencing を利用した本解析法により、患者血漿中のウイルス RNA ゲノム配列を決定し、RF と重複感染症例のウイルス混合とを効率的かつ簡便に区別することができた。遺伝子断片クローニングや限界希釈が前提となる従来の配列決定法とは異なり、Nanopore sequencing により HIV-1 近全長ゲノムを決定することが可能である。今後、HIV-1 の遺伝的多様性が拡大する中で本解析法は分子疫学調査や薬剤耐性検査において有益な手段として活用されることが期待される。

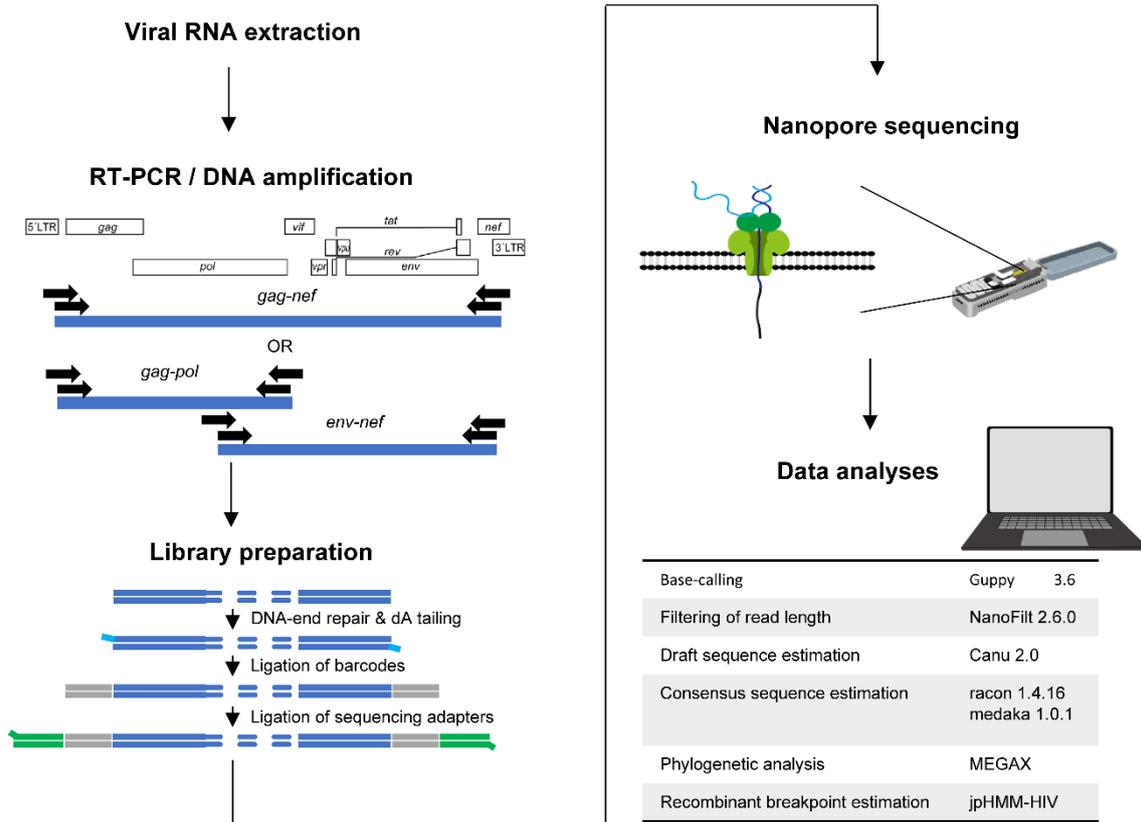


Fig. 1 A flowchart of the nanopore sequencing protocol applied to the near-full-length HIV-1 genome in this study.

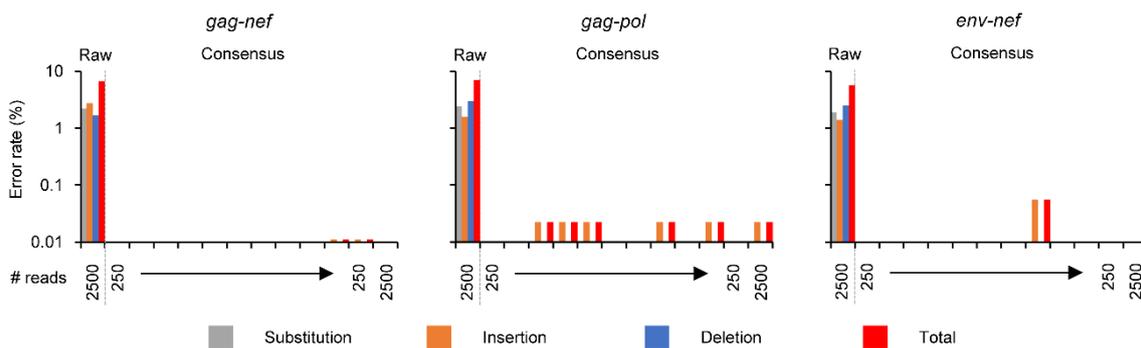


Fig. 2 Error rate (%) among raw reads and consensus sequences generated by nanopore sequencing. The proviral DNA, pNL4-3, was used for analysis. Individual error patterns (substitutions, insertions, deletions, and their total) are colored. Consensus sequences were estimated repeatedly from 10 different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

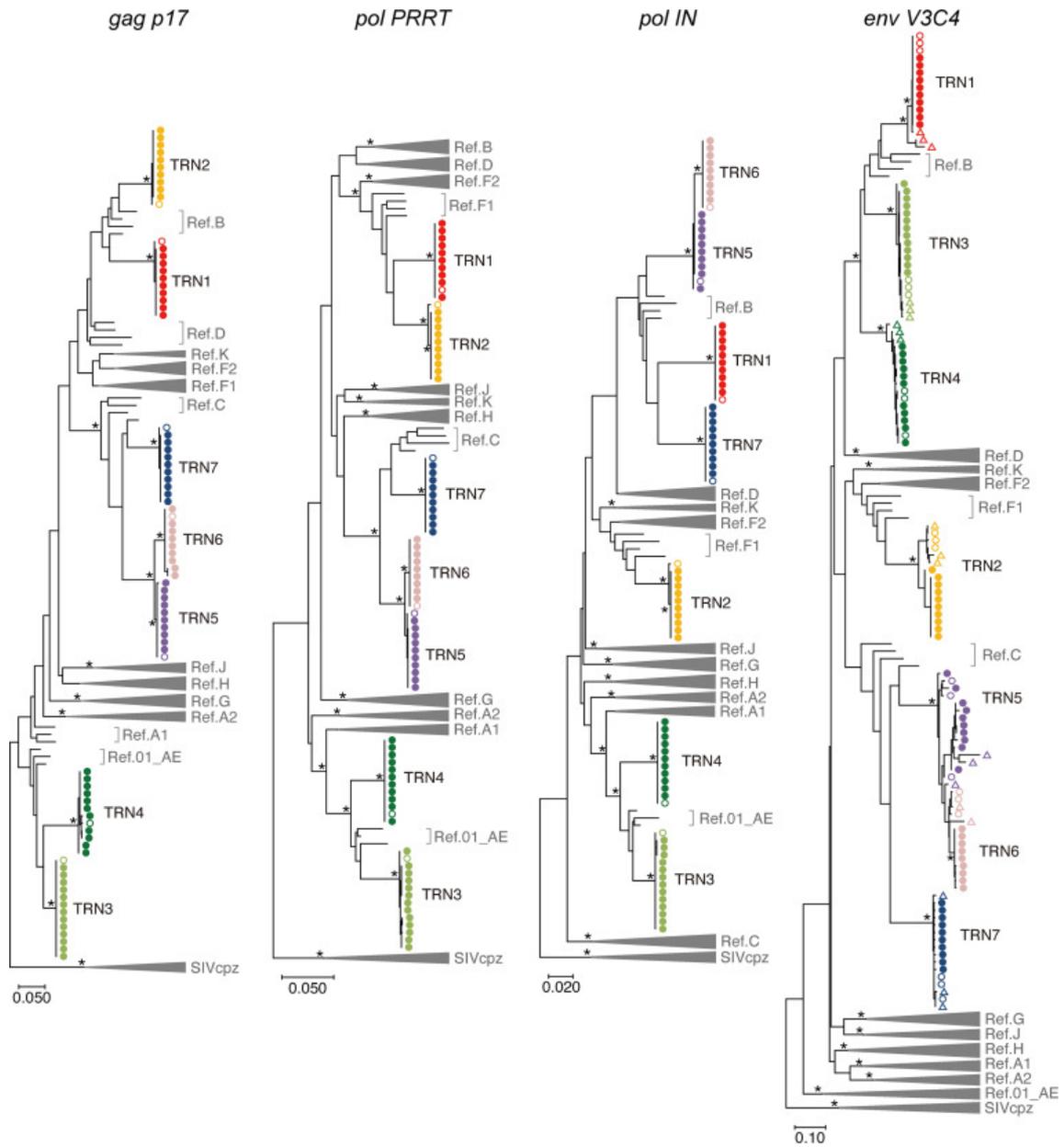


Fig. 3 Phylogenies of the RF sequences identified by nanopore sequencing and Sanger sequencing. Maximum likelihood trees with the GTR model are shown for four regions: *gag p17* (positions 790 to 1185 of HXB2), *pol PRRT* (2253 to 3269), *pol IN* (4230 to 5093), and *env V3C4* (7114 to 7589). The reference sequences (Ref.) of HIV-1 major subtypes (A1, A2, B, C, D, F1, F2, G, H, J, K, and CRF01_AE) and three SIVcpz strains are indicated in the tree. The three SIVcpz sequences are used as outliers (GenBank no. DQ373064, DQ373063, and EF535994 for SIVcpzLB7, SIVcpzMB66, and SIVcpzMB897, respectively). Bootstrap values were calculated by 500 replicates. Branches with bootstrap values of at least 0.95 are highlighted with asterisks. The viral full-genome sequences (closed circles) identified by nanopore sequencing and the viral RNA (open circles) and proviral DNA sequences (triangles) identified by Sanger sequencing are indicated.

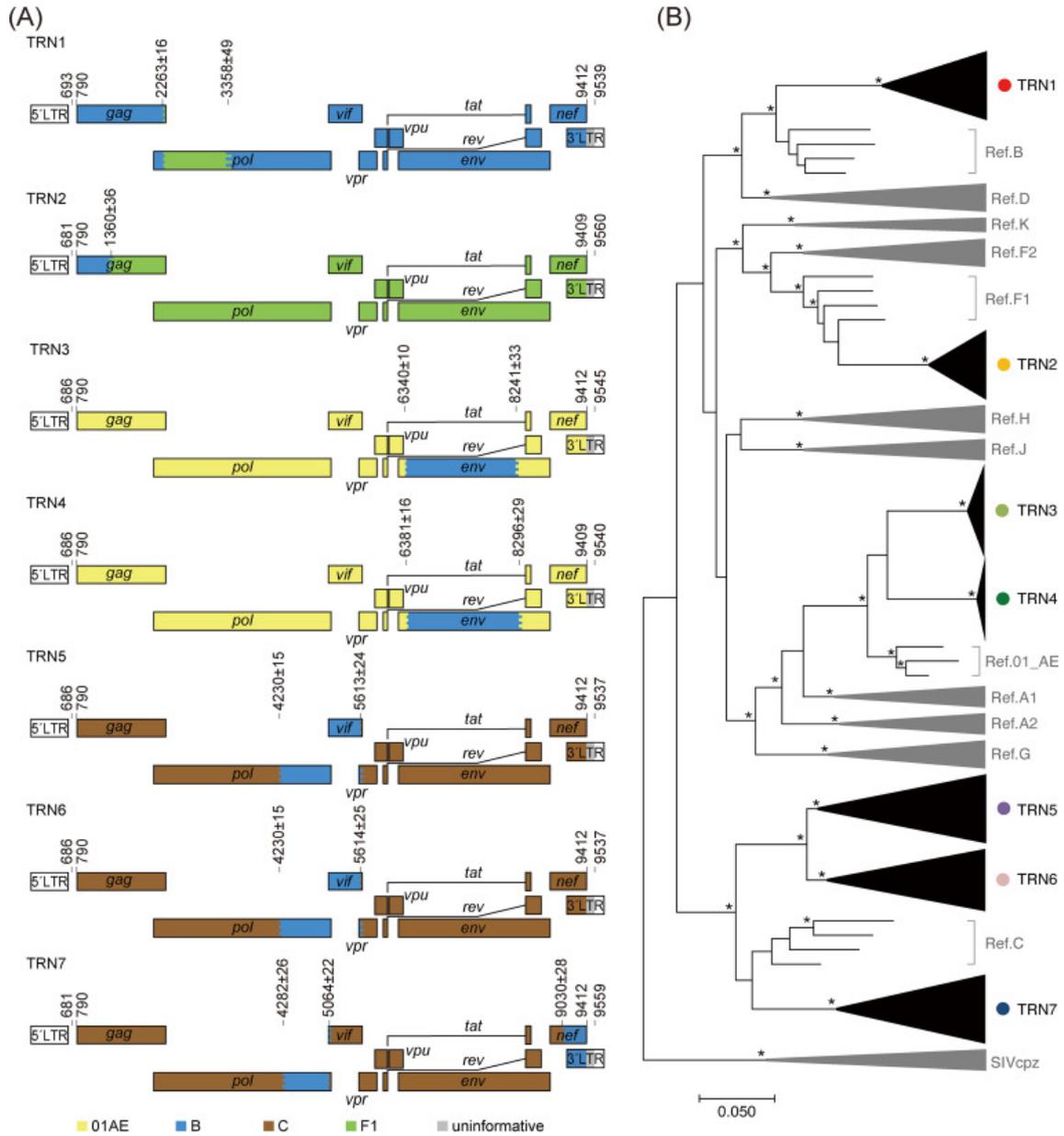


Fig. 4 Characterization of the HIV-1 RFs. (A) Genome maps of the RFs. The numbers on each map denote the nucleotide positions of the breakpoints according to the HXB2 reference numbering system. The colored stripes in the maps indicate the intervals where the recombination breakpoints are estimated to be located. (B) Maximum likelihood tree for pangenomic consensus sequences obtained by nanopore sequencing and complete genomes of the major subtype references. The three sequences of SIVcpz are used as outliers. Branches with bootstrap values (based on 500 replicates) of at least 0.95 are shown with asterisks.

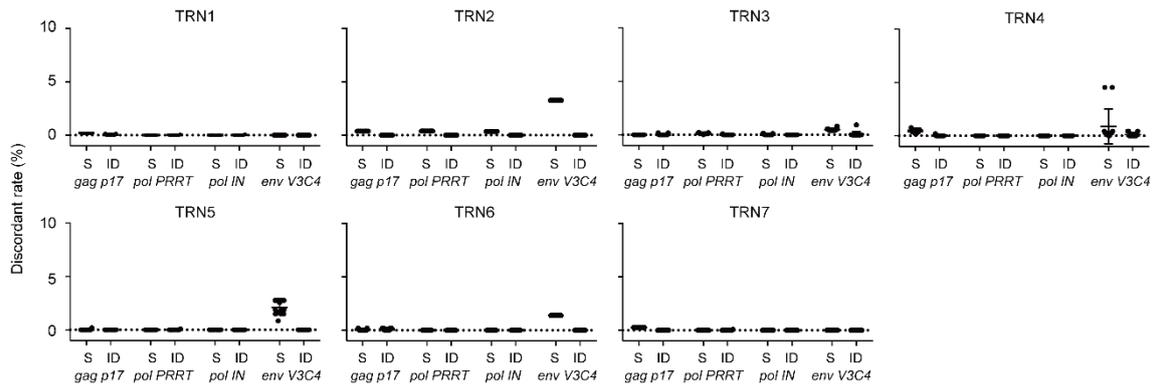


Fig. 5 Ratio of mismatched bases between Sanger and nanopore sequencing. The viral genome sequences in the gag p17 (positions 790 to 1,185 relative to the reference HXB2), pol PRRT (2,253 to 3,269), pol IN (4,230 to 5,093), and env V3C4 (7,114 to 7,589) regions were compared. The ratios at the nucleotide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

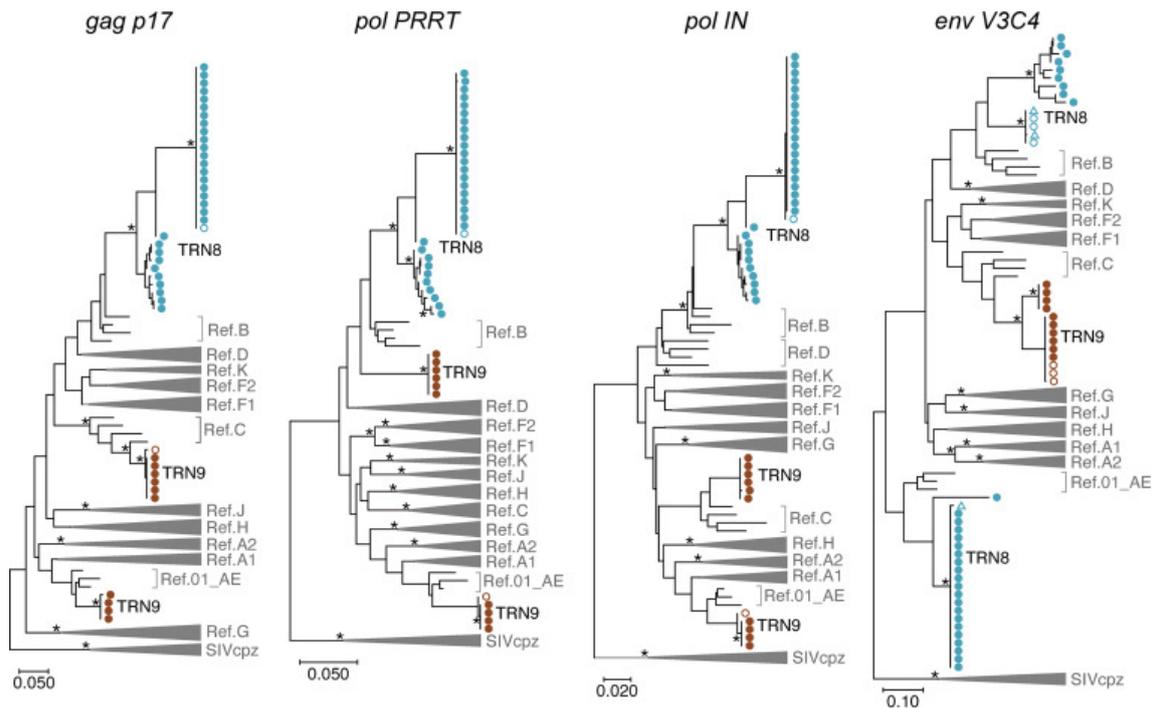


Fig. 6 Phylogenies of consensus sequences determined by nanopore sequencing and Sanger methods. Maximum likelihood trees generated with the GTR model are shown for four regions, *gag p17*, *pol PRRT*, *pol IN*, and *env V3C4*, with the nucleotide numbers corresponding to those indicated in Fig. 3. The reference sequences of major HIV-1 subtypes (A1, A2, B, C, D, F1, F2, G, H, J, K, and CRF01_AE) and three SIVcpz strains (LB7, MB66, and MB897) are shown in the trees. Branches with bootstrap values (based on 500 replicates) of at least 0.95 are indicated with asterisks. The viral RNA sequences obtained by Sanger sequencing and nanopore sequencing are indicated with open and closed circles, respectively. The proviral DNA sequences generated by Sanger sequencing are shown with triangles.

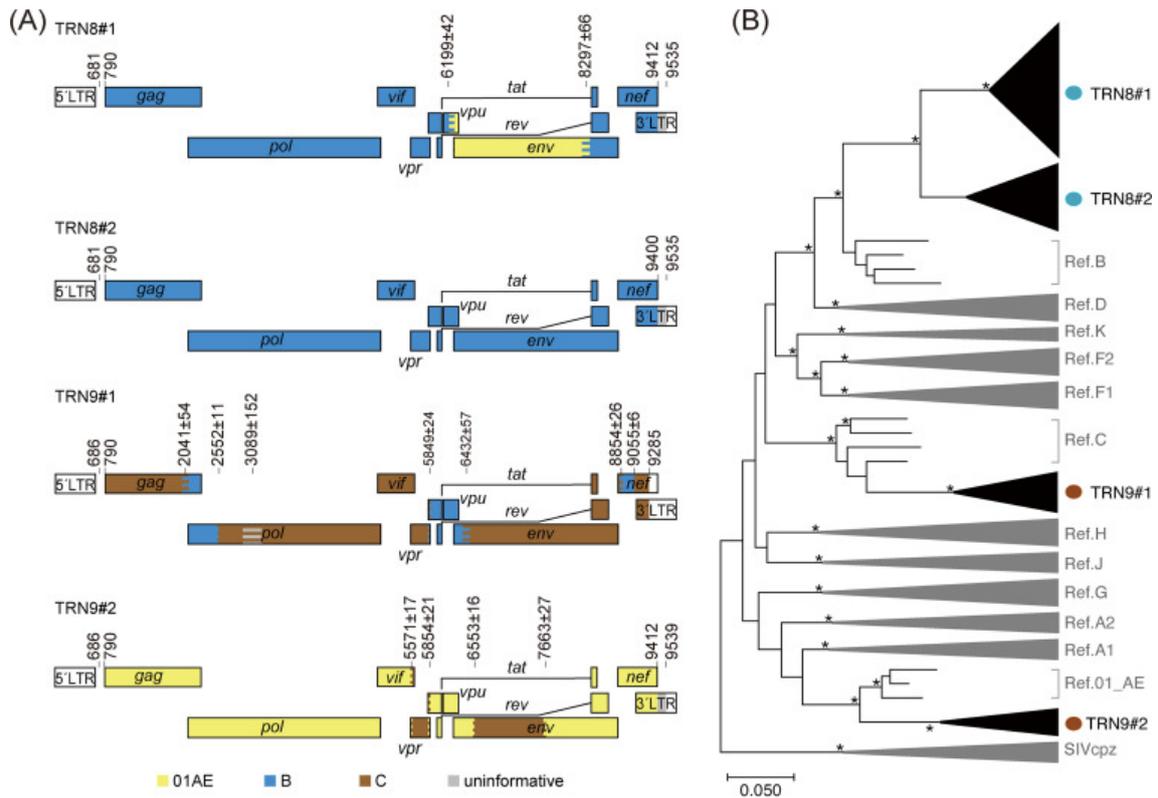


Fig. 7 Identification of dual-HIV-1 infection cases. (A) Genome maps of representative sequences obtained by nanopore sequencing. Two different patterns of the genome sequences were identified as dual-infection cases in TRN8 and TRN9. The numbers on each map represent the breakpoints according to the HXB2 reference numbering system. The colored stripes in the maps indicate the intervals where the recombination breakpoints are estimated to be located. (B) Maximum likelihood tree for pangenomic consensus sequences determined by nanopore sequencing and complete genomes of major HIV-1 references. Three sequences of SIVcpz (LB7, MB66, and MB897) are used as outliers. Branches with bootstrap values (based on 500 replicates) of at least 0.95 are indicated with asterisks.