Journal of Virological Methods 274 (2019) 113737

Contents lists available at ScienceDirect

Journal of Virological Methods

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

Single genome sequencing of near full-length HIV-1 RNA using a limiting dilution approach

Laura Hebberecht, Leen Vancoillie, Marlies Schauvliege, Delfien Staelens, Els Demecheleer, Jarryt Hardy, Virginie Mortier, Chris Verhofstede*

Aids Reference Laboratory, Department of Diagnostic Sciences, Ghent University, Ghent, Belgium

ARTICLE INFO

Keywords:
Near full-length HIV-1 sequencing
Single genome sequencing
HIV-1 RNA limiting dilution
Recombination

ABSTRACT

Sequencing very long stretches of the HIV-1 genome can advance studies on virus evolution and *in vivo* recombination but remains technically challenging. We developed an efficient procedure to sequence near full-length HIV-1 RNA using a two-amplicon approach. The whole genome was successfully amplified for 107 (88%) of 121 plasma samples including samples from patients infected with HIV-1 subtype A1, B, C, D, F1, G, H, CRF01_AE and CRF02_AG. For the 17 samples with a viral load below 1000 c/ml and the 104 samples with a viral load above 1000 c/ml, the amplification efficiency was respectively 53% and 94%. The sensitivity of the method was further evaluated using limiting dilution of RNA extracted from a plasma pool containing an equimolar mixture of three HIV-1 subtypes (B, C and CRF02_AG) and diluted before and after cDNA generation. Both RNA and cDNA dilution showed comparable sensitivity and equal accuracy in reflecting the subtype distribution of the plasma pool. One single event of *in vitro* recombination was detected amongst the 41 sequences obtained after cDNA dilution but no indications for *in vitro* recombination were found after RNA dilution. In conclusion, a two-amplicon strategy and limiting dilution of viral RNA followed by reverse transcription, nested PCR and Sanger sequencing, allows near full genome sequencing of individual HIV-1 RNA molecules. This method will be a valuable tool in the study of virus evolution and recombination.

1. Introduction

The replicative cycle of HIV is characterized by high rates of nucleotide misincorporations, insertions, deletions and recombinations (Alizon and Montagnier, 1987; Cromer et al., 2016; Cuevas et al., 2015). As a result, the homogeneous virus population typically present early after infection soon evolves to a quasispecies of genetic variants (Domingo et al., 1985). Characterization of this quasispecies can provide key information on the interplay between virus and host and on virus evolution in general. Most studies on virus evolution are restricted to relatively short fragments of the HIV-1 genome, often located in the fast evolving *envelope* or *gag* region. Although next-generation sequencing (NGS) technologies offer new opportunities for full genome sequencing, an important constraint of NGS is the short read length of the sequences and the resulting difficulty to link sequences over longer distance. The more recently developed NGS technologies like PacBio RS or Oxford Nanopore MinION may overcome these shortcomings but they still suffer from high error rates (Dilernia et al., 2015).

Both the traditional Sanger sequencing and NGS require initial amplification of the genomic region of interest. Amplification of the HIV-1 RNA extracted from blood plasma as a single strand is extremely challenging. As far as we know, only one successful attempt has been described but this method was only evaluated on HIV-1 subtype C (Rousseau et al., 2006). There were a few attempts to cover the genome with two amplicons (Aralaguppe et al., 2016; Grossmann et al., 2015; Nadai et al., 2008; Salazar-Gonzalez et al., 2009; Tovanabutra et al., 2010) but in the majority of studies a four to six amplicon strategy is used (Bimber et al., 2010; Brener et al., 2015; Di Giallonardo et al., 2013; Fang et al., 1999; Gall et al., 2012; Ode et al., 2015; Walker et al., 2012; Zagordi et al., 2010; Zanini et al., 2015). Reliable linking of the sequences obtained from four or six independent amplicons is nearly impossible, especially when analyzing highly divergent virus populations. For NGS there is the additional problem that the multiple sequences generated for each amplicon are of small read lengths. The accurate assembly of these short sequences to reconstruct the individual amplicons is also challenging (Posada-Cespedes et al., 2017). All these problems impede research aimed at characterizing the HIV-1 quasispecies composition, tracking immune evasion or studying *in vivo*

* Corresponding author.
E-mail address: chris.verhofstede@ugent.be (C. Verhofstede).

https://doi.org/10.1016/j.jviromet.2019.113737
Received 28 March 2019; Received in revised form 20 September 2019; Accepted 22 September 2019
Available online 25 September 2019

recombination. As we were particularly interested in the study of recombination, we sought a way to amplify and sequence individual viral variants over an as long as possible length. We developed a PCR protocol that covered the near full-length HIV genome with two amplicons. We then assessed the sensitivity of the method and the possibility to execute the amplification on single viral RNA molecules using a limiting dilution approach. With limiting dilution, in essence, the viral genetic material is diluted to a point where no more than 30% of subsequently performed PCR reactions are positive. In optimal conditions, the mathematical probability that the amplicons subsequently generated originate from a single genome, is higher than 80% (Rodrigo et al., 1997). Limiting dilution was and still is an often applied method for individual variant sequencing. Individual amplification of single variants is the only way to prevent PCR-induced recombination, a frequently observed phenomenon when performing bulk sequencing (Cronn et al., 2002; Gorzer et al., 2010; Judo et al., 1998; Liu et al., 2014; Meyerhans et al., 1990; Yang et al., 1996). When using viral RNA as input material for individual variant sequencing, it is standard practice to first reverse transcribe the RNA to cDNA followed by cDNA end point dilution before amplification. *In vitro* reverse transcription however is a very inefficient reaction, prone to recombination errors and founder effects. The final result may be a library of amplicons that does not adequately reflect the variant composition of the original sample (Miranda and Steward, 2017; Salazar-Gonzalez et al., 2008; Svarovskaia et al., 2003).

In the search for a more reliable method to study *in vivo* recombination of HIV-1 in patients co- or super-infected with different HIV-1 variants we developed an amplification and sequencing strategy that covers the near full-length HIV-1 genome with two amplicons. The amplification method was validated on a selection of plasma samples with a broad range of viral loads and HIV-1 subtypes and on limiting dilutions of both RNA and cDNA from a plasma pool with 3 different HIV-1 subtypes.

2. Materials and methods

2.1. Ethical approval

The study was approved by the Ethical Committee of Ghent University Hospital (reference number 2014/0173). Patients provided signed informed consent for the use of their leftover plasma for scientific research. All samples were anonymized before processing.

2.2. Sample selection

Plasma samples were collected from 121 HIV-1 infected individuals. For 87 individuals the infection stage at the time of sampling was defined, indicating collection during acute infection in 66 and during chronic infection in 21. Patients were considered acutely infected if they presented with detectable HIV p24 antigen and viral RNA in the blood in the absence of HIV antibodies or if they were classified as recently infected after testing in the Sedia HIV-1 LAg Avidity EIA test (Sedia Biosciences, Portland, USA). This test has an estimated recency window of 3 to 4 months (Verhofstede et al., 2017).

The selected samples covered a wide range of viral loads (mean: 58,884 copies(c)/ml) and included 17 samples with a viral load of < 1000 c/ml; 13 samples with a viral load between 1000 and 10,000 c/ml and 91 samples with a viral load of > 10,000 c/ml. The subtype variation reflected the subtype distribution in our country including subtype B (n = 81), A1 (n = 5), C (n = 3), F1 (n = 8) and CRF02_AG (n = 13) and some representatives of less frequently observed subtypes A6 (n = 2), D (n = 2), G (n = 2), H (n = 1), CRF01_AE (n = 3) and CRF06_cpx (n = 1).

2.3. Plasma pool

A plasma pool was prepared using plasma collected before seroconversion from three patients infected with genetically different viruses (respectively HIV-1 subtype B, C and CRF02_AG). Collection during the pre-seroconversion stage ensured low intra-patient genetic variability of the virus. Using the HIV-1 viral load as a reference, the 3 samples were diluted in HIV negative plasma to a calculated concentration of 50,000 RNA c/ml each. Re-analyses of the viral load of the diluted samples showed a result of 44,000 c/ml, 55,000 c/ml and 63,500 c/ml respectively, for the subtype B, C and CRF02_AG sample. Equal volumes of each of these dilutions were pooled to create a pool with nearly equimolar representation of the three subtypes.

2.4. RNA extraction

Viral RNA was extracted using the QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. For each extraction, 140 µl of plasma was used as input material. Purified RNA was eluted in 60 µl and stored at −20 °C until use.

For the plasma pool, 20 individual extractions were performed in order to obtain sufficient amounts of RNA for all experiments and to minimize the impact of extraction-induced differences in variant composition. The resulting extracts were pooled, aliquoted and stored at −20 °C until use.

2.5. Reverse transcription

Bulk reverse transcription (RT) was done with SuperScript™ III (Thermo Fisher Scientific, Massachusets, USA) and the reverse primers F3_rev1 and F6_rev1 (Table 1) to reverse transcribe respectively the 5′ and the 3′ half of the HIV-1 genome. For each reaction, 2.5 mM reverse primer (1 µl) and 10 mM dNTP's (1 µl) was added to 11 µl RNA. After heating the mixture to 65 °C for 5 min and cooling down for 1 min, a reaction mix of 5X RT buffer (4 µl), 100 mM DTT (1 µl), 40 U RNAsin (1 µl) and 200 U Superscript RT (1 µl) was added and the reaction mixture was heated to 50 °C for 60 min and 70 °C for 15 min followed by a cool down to 4 °C. The resulting cDNA was then amplified using the Platinum™ Taq DNA Polymerase High Fidelity kit (Thermo Fisher Scientific, Massachusetts, USA). The primer sets used for amplification consisted of the same primers as used for cDNA generation supplemented with the sense primer F1_fw1 for the 5′ fragment and F4_fw1 for the 3′ fragment. For each reaction, 5 µl of the generated cDNA was added to 45 µl of a reaction mix containing molecular biology grade water (34.8 µl), 10X High Fidelity buffer (5 µl), 50 mM $MgSO_4$ (2 µl), 10 mM dNTPs (1 µl), 1 U Platinum Taq (0.2 µl) and 10 pmol/µl primer (1 µl of each). Cycling conditions were 15 s at 94 °C, 30 s at 50 °C and 5 min at 68 °C for 30 repetitions with a final extension of 5 min at 68 °C and a cool down to 4 °C.

For the reverse transcription of RNA after limiting dilution, the reverse transcription step and first round PCR were combined in a single reaction using the Superscript III One-Step RT-PCR Platinum Taq HiFi kit™ (Thermo Fisher, Massachusets, USA). This kit contains an enzyme mixture of SuperScript™ III and Platinum™ Taq DNA Polymerase. Fourteen µl of RNA was added to a reaction mix containing molecular biology grade water (8 µl), 2X reaction mix (25 µl), enzyme mix (1 µl) and 10 pmol/µl (1 µl) of either the primers defining the 5′ part of the genome (F1_fw1 and F3_rev1) or the primers defining the 3′ part of the genome (F4_fw1 and F6_rev1). The reaction conditions were 30 min at 50 °C followed by 2 min of denaturation at 94 °C for the reverse transcription followed by 30 amplification cycles (15 s at 94 °C, 30 s at 50 °C and 5 min at 68 °C). The reaction was ended with a final extension of 5 min at 68 °C followed by a cool down to 4 °C.

Table 1
Primers used for amplification and sequencing of the near full-length HIV-1 genome. RT = reverse transcription; Seq = sequencing; *Position based on HXB2 numbering; ˚Alternative primers.

Primer	Direction	Start position*	Stop position*	Sequence (5'→'3)	Function
F1_fw1	Forward	524	547	CTCAATAAAGCTTGCCTTGAGTGC	First PCR
F3_rev1	Reverse	5185	5213	GGGATGTGTACTTCTGAACTTAYTYTTGG	RT + First PCR
F4_fw1	Forward	4899	4922	CGGGTTTATTWCAGRGAGAGCAGA	First PCR
F6_rev1	Reverse	9592	9614	ATTGAGGCTTAAGCAGTGGGTTC	RT + First PCR
F1_fw2	Forward	551	571	AAGTAGTGTGTGCCCGTCTGT	Nested PCR + Seq
F3_rev2	Reverse	5040	5061	CACCTGCCATCTGTTTTCCATA	Nested PCR + Seq
F4_fw2	Forward	4956	4974	TGGAAAGGTGAAGGGGCAG	Nested PCR + Seq
F6_rev2	Reverse	9567	9582	CCAGAGAGCTCCCAGG	Nested PCR + Seq
G40	Forward	1075	1093	GACACCAAGGAAGCTTTAG	Seq
GAG8	Reverse	1473	1500	TGCTATGTCACTTCCCCTTTGTTCTCTC	Seq
1231F	Forward	1231	1255	TCACCTAGAACTTTRAATGCATGGG	Seq
1817F	Forward	1817	1834	TAGAAGAAATGATGACAG	Seq
F2_fw2	Forward	2022	2039	GGGCTGTTGGARATGTGG	Seq
PR3	Forward	2148	2166	AGAGCCAACAGCCCCACCA	Seq
RT2955	Reverse	2955	2977	CTAATYCCTGGYGTYTCATTRTT	Seq
F1_rev2	Reverse	2252	2272	TGCCAAAGAGTGATYTGAGGG	Seq
R5	Reverse	3492	3511	GGGTCATAATACACTCCATG	Seq
IN_MW1	Forward	2997	3019	CCACARGGATGGAAAGGATCACC	Seq
F3_fw2	Forward	3681	3703	GAAAGCATAGTRATATGGGGAAA	Seq
F2_rev2	Reverse	3776	3799	ACAAACTCCCAYTCAGGAATCCA	Seq
4350R	Reverse	4326	4350	CACAGCTAGCTACTATTTCTTTTGC	Seq
IN99	Reverse	4482	4504	TCTGCTGGRATRACYTCTGCYTC	Seq
IN22	Forward	4074	4095	CAACCAGAYAARAGTGARTCAG	Seq
IN88	Forward	4608	4627	AAGGCIGCMTGYTGGTGGGC	Seq
ENVoutF1	Forward	5550	5574	AGARGAYAGATGGAACAAGCCCCAG	Seq
5861F	Forward	5861	5884	TGGAAGCATCCRGGAAGTCAGCCT	Seq
ACC8R	Reverse	5968	5989	TCTCCGCTTCTTCCTGCCATAG	Seq
F4_rev2	Reverse	6429	6450	GTACACAGGCATGTGTRGCCCA	Seq
F5_rev2	Reverse	7992	8012	CCAAATYCCYAGGAGCTGTTG	Seq
7382	Reverse	7357	7382	ATTACARTAGAAAAATTCYCCTCYAC	Seq
7336	Reverse	7318	7336	ATTTCTGGRTCYCCKCCTG	Seq
6990	Forward	6990	7012	TCAACHCAAYTRCTGTTAAATGG	Seq
6951	Forward	6951	6973	AGYRCAGTACAATGYACACATGG	Seq
7373R	Reverse	7350	7373	GAAAAATTCTCCTCYACAATTAAA	Seq
7761F	Forward	7761	7787	GTGGGAATAGGAGCTGTGTTCCTTGGG	Seq
KVL072	Forward	8330	8355	AATAGAGTTAGGMAGGGATACTCACC	Seq
8445R	Reverse	8424	8445	CTCTCTCTCCACCTTCTTCTTC	Seq
ENVM	Reverse	9058	9086	TAGCCCTTCCAGTCCCCCCTTTTCTTTTA	Seq
F_ENVM	Forward	9058	9086	TAAAAGAAAAGGGGGGACTGGAAGGGCTA	Seq
nefyn05	Reverse	9157	9181	GTGTGTAGTTCTGCCAATCAGGGAA	Seq
˚ENV6889R	Reverse	6865	6889	CCAGCYGGRGCACAATAATGTATRG	Seq
˚ENV6865F	Forward	6865	6889	CYATACATTATTGTGCYCCRGCTGG	Seq
˚ENV7538R	Reverse	7521	7538	RGGAGGRGCATAYATTGC	Seq
˚ENV7521F	Forward	7521	7538	GCAATRTATGCYCCTCCY	Seq
˚ENV7957R	Reverse	7936	7957	GCCTGGAGCTGYTTRATGCCCC	Seq
˚ENV7936F	Forward	7936	7957	GGGGCATYAARCAGCTCCAGGC	Seq
˚ENV6328F	Forward	6328	6348	GGGTYACAGTHTATTATGGRG	Seq

2.6. Nested PCR

Two and a half µl of the amplicons generated during first-round amplification was transferred to 47.5 µl of a reaction mix containing 1 U (0.2 µl) Platinum™ Taq DNA Polymerase (Thermo Fisher Scientific, Massachusetts, USA), 10 mM dNTPs (1 µl), 10X High Fidelity buffer (5 µl), 50 mM $MgSO_4$ (2 µl), molecular biology grade water (37.3 µl) and 10 pmol/µl (1 µl) of the nested primers F1_fw2 and F3_rev2 for the 5′ fragment or F4_fw2 and F6_rev2 for the 3′ fragment. The reaction mixes (total volume of 50 µl for each fragment) were then subjected to 30 cycles of 30 s at 94 ˚C, 15 s at 55 ˚C and 4 min at 60 ˚C. The presence of an amplicon with a respective length of 4511 nucleotides (nt) for the 5′ fragment and 4627 nt for the 3′ fragment was visualized on a 2% agarose gel (Roche Applied Science, Mannheim, Germany), using the GeneRuler 1 kb Plus DNA ladder as DNA size reference (Thermo Fisher Scientific, Massachusetts, USA).

2.7. Sanger sequencing

Positive PCR products were purified with Agencourt AMPure beads (Beckman Coulter, California, USA). For BigDye™ Terminator v3.1 Cycle Sequencing (Life Technologies, Carlsbad CA, USA), 2 µl of the purified amplicon was added to a mixture of 5X Sequencing Buffer (1.75 µl), 1X BigDye® Terminator v3.1 Ready Reaction Mix (0.5 µl), 2 pmol/µl primer (2.5 µl) and molecular biology grade water (3.25 µl). For each amplicon at least 18 sequencing reactions were run with the sequencing primers listed in Table 1. Reactions were subjected to 30 cycles of 15 s at 94 ˚C, 30 s at 50 ˚C and 5 s at 68 ˚C, followed by a cool down to 4 ˚C and then purified using Agencourt CleanSEQ™ beads (Beckman Coulter, California, USA) before analysis on the 3500 or 3500 XL Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, Massachusetts, USA). Proofreading of the chromatograms was done with Smartgene IDNS (Smartgene, Zug, Switzerland). Amplicons with mixed nucleotides were omitted from subsequent analyses.

Because we encountered problems to sequence the highly variable *envelope* region of some HIV-1 variants, additional primers were designed to improve the sequencing of this region. These primers are listed in Table 1 as 'alternative' primers. They were validated on a selection of 3′ amplicons from 6 patients, including the three patients that constituted the plasma pool. Due to shortage of material however the

Table 2
Subtype distribution of the amplicons generated from a plasma pool containing an equimolar mix of HIV-1 subtype B, C and CRF02_AG after limiting dilution of respectively RNA and cDNA.

	Limiting dilution				
	RNA		**cDNA**		
	5' amplicon	3' amplicon	5' amplicon	3' amplicon	Total
Number of clones	**28**	**25**	**21**	**20**	**94**
Subtype CRF02_AG	15 (53.6%)	8 (32.0%)	9 (42.9%)	3 (15.0%)	35 (37.2%)
Subtype C	10 (35.7%)	7 (28.0%)	8 (38.1%)	6 (30.0%)	31 (33.0%)
Subtype B	3 (10.7%)	10 (40.0%)	4 (19.0%)	10 (50.0%)	27 (28.7%)
Recombinant form	–	–	–	1 (5.0%)	1 (1.1%)

primers could not be tested on the different individual amplicons generated from the plasma pool.

2.8. Viral load

Viral load analysis was performed using the Cobas ® HIV-1 test (Roche Diagnostics, Basel, Switzerland).

2.9. Subtyping and phylogenetic analysis

Subtyping was done based on the *polymerase* sequences obtained through standard of care baseline resistance testing using the HIV-1 subtyping tool COMET (COntext-based Modeling for Expeditious Typing)(Struck et al., 2014). JpHMM (jumping profile Hidden Markov Model) (Schultz et al., 2012) was used to visualize potential recombination events. Identified recombinants were confirmed using the bootscanning method in Simplot (Lole et al., 1999).

Phylogenetic trees were constructed with PhyML 3.0 using the AIC (Akaike Information Criterion) selection, the NNI (Nearest Neighbour Interchange) tree-rearrangement and the aLRT SH-like (approximate Likelihood Ratio test and Shimodaira–Hasegawa) method for branch support (Guindon et al., 2010). Trees were visualized with iTol (Letunic and Bork, 2016).

The Genbank accession numbers of the sequences generated during this study are: MN485971-MN486047.

3. Results

3.1. Amplification of near full-length HIV-1 RNA

Primers for reverse transcription, first round and nested PCR were selected from the work of Zanini et al. (2015). For their amplicon library preparation, these authors generated six overlapping amplicons, together covering nt 571 to 9567 of the HIV genome. In an attempt to reduce the number of independent amplicons, we first tried to combine the outermost 5′ and 3′ primers but all attempts to generate a single full-length amplicon failed. We then combined the outermost 5′ and 3′ primers (F1_fw1 and F6_rev1) with primers located in the center of the genome (F3_rev1 and F4_fw1) for the reverse transcription and first round PCR and the primers F1_fw2 and F3_rev2 and F4_fw2 and F6_rev2 for the nested amplification. This allowed to generate two amplicons with a short overlap, covering respectively nt 551 to 5061 (5′ fragment) and nt 4899 to 9582 (3′ fragment) of the HIV genome. The success rate and sensitivity of the reverse transcription PCR for 121 plasma samples was 90% for the 5′ fragment and 95% for the 3′ fragment. For 88% both amplicons could be generated. For the 17 samples with a viral load of less than 1000 c/ml that were included, the success rate was 65% for the 5′ fragment and 76% for the 3′ fragment. For 9 samples (53%) both fragments were generated. The viral loads of these 9 samples were 144 c/ml, 288 c/ml, 324 c/ml, 330 c/ml, 389 c/ml, 449 c/ml, 492 c/ml, 575 c/ml and 610 c/ml.

3.2. RNA limiting dilution

To further assess the sensitivity of the two-amplicon reverse transcription PCR, serial dilutions of viral RNA extracted from a plasma pool with subtype B, C, and CRF02_AG viruses were tested. RNA was diluted 1/1, 1/5, 1/10, 1/20, 1/30, 1/40, 1/50, 1/60 in molecular biology grade water (VWR, Pennsylvania, USA) and each dilution was subjected to reverse transcription PCR. Successful amplification was seen up to a dilution of 1/40 and 1/50 respectively for the 5′ and 3′ amplicon. These dilutions were selected as probable endpoints and 10 replicates of both were retested. With 3/10 positive reactions for the 1/50 dilution, for both fragments, this dilution, that corresponded to a calculated input equivalent of 32.8 RNA copies per reaction, complied with the conditions for being considered as the endpoint. To increase the number of individual amplicons for subsequent sequencing, 90 replicates of the 1/50 dilution were additionally subjected to RT-PCR. Finally, respectively 28 and 25 individual 5′ amplicons and 3′ amplicons were available for sequencing. The subtype distribution of these amplicons is shown in Table 2. No intersubtype recombinants were detected.

3.3. cDNA limiting dilution

cDNA resulting from bulk reverse transcription of plasma pool RNA, was diluted 1/1, 1/5, 1/10, 1/20 and 1/40 in molecular biology grade water (VWR, Pennsylvania, USA) and subjected to two rounds of PCR to generate the 5′ and 3′ amplicons. Positive reactions were obtained up to the 1/10 dilution for the 5′ fragment and up to the 1/5 dilution for the 3′ fragment. For both dilutions the reactions were repeated in 10-fold. This revealed a 30% success for the 1/5 dilution for both fragments. This dilution, corresponding to a calculated input equivalent of 64.3 copies viral RNA per reaction, was considered as the endpoint and 90 additional replicates were tested to finally result in 21 and 20 individual amplicons of respectively the 5′ and the 3′ fragment. The subtype distribution of the resulting sequences is shown in Table 2. One of the 3′ amplicons (cDNA_F2_8) showed clear indications of *in vitro* recombination, with a small stretch of subtype C in a CRF02_AG background. The recombination is demonstrated by both the highlighter plot (Fig. 1B), the phylogenetic tree (Fig. 2B) and Simplot (**S1**).

3.4. Sequencing reactions

Amplicons were sequenced bidirectional using 36 sequencing primers including the primers used for the nested PCR. For the 3′ amplicons, some problems were encountered for part of the variable *envelope* gene with failing sequencing reactions for the region between nt 6391 and 6723 for subtype C virus and between nt 7965 and 8331 for CRF02_AG (Fig. 1B). An overview of all sequences presented as highlighter plots with the HXB2 sequence as a reference is shown in Fig. 1A for the 5′ amplicons, and in Fig. 1B for the 3′ amplicons. The phylogenetic trees are illustrated respectively in Fig. 2A and B.

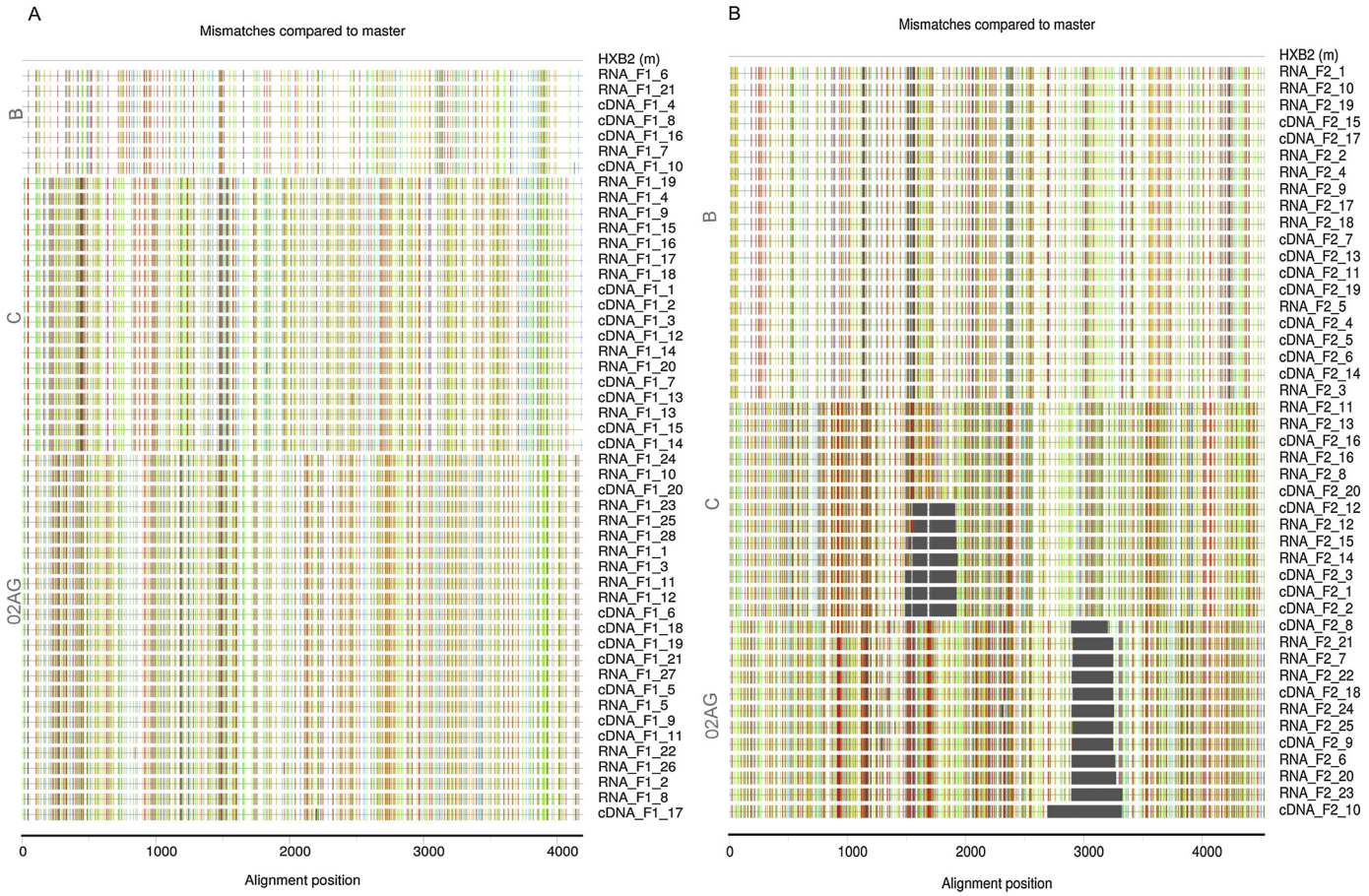

Fig. 1. Highlighter plot of sequences generated after limiting dilution of HIV-1 RNA and cDNA from a plasma pool with 3 HIV-1 variants.
(A) Highlighter plot of the 5′ amplicon sequences. (B) Highlighter plot of the 3′ amplicon sequences. Sequences are aligned to the HXB2 reference sequence (master sequence (m)). Mutations are color coded. Zones of bad sequence quality are displayed in gray. The corresponding subtype is indicated on the left. The sequences are identified with the target of limiting dilution (RNA or cDNA), the fragment (F1: 5′ amplicon; F2: 3′ amplicon) and a serial number.

4. Discussion

The aim of this study was to develop a method for amplification and sequencing of nearly complete individual HIV-1 viral RNA molecules, using a minimum number of amplified fragments. A two-amplicon approach, generating amplicons with a length of 4511 nt and 4627 nt, respectively covering the 5′ and 3′ half of the viral genome, was designed and validated. Amplification primers were selected from the study of Zanini et al. (2015). Successful amplification of both fragments was obtained for viral RNA extracted from the plasma of 107 (88%) of 121 HIV-1 infected individuals, including 9 (53%) of 17 samples with a viral load of < 1000 c/ml. Aralaguppe et al. (2016) and Grossmann et al. (2015) also successfully used a two-amplicon approach to sequence the whole HIV-1 genome but the sensitivity of their protocol was limited to 3000 c/ml. In other studies, the whole viral genome was reconstructed with four to six overlapping amplicons with read lengths of less than 4000 nt (Alampalli et al., 2017; Berg et al., 2016; Brener et al., 2015; Di Giallonardo et al., 2013; Dukhovlinova et al., 2018; Gall et al., 2012; Ode et al., 2015; Walker et al., 2012; Zanini et al., 2015). The problem of this more extensive fragmentation is that genomic linkage is limited to the individual pieces. This hampers insight on the overall evolution of the genome and is a major drawback in studies on *in vivo* recombination.

The primers used were originally designed to cover a broad range of viral genetic variability (Zanini et al., 2015). The high specificity for different subtypes was confirmed by the successful amplification of subtype A1, A6, B, C, D, F1, G, H, CRF01_AE and CRF02_AG.

Because of the high intra-patient variability of HIV-1, studying the genetic evolution not only requires the generation of long amplicons but also the possibility to sequence individual viruses. NGS seems to be the most appropriate method to do so but up to now it remains extremely difficult to reconstruct long sequences from the short read length sequences generated by NGS. We therefore explored the potential of limiting dilution sequencing using either RNA or cDNA as starting material and were surprised to see that despite the long amplification length, both are feasible and equally sensitive. *In vitro* reverse transcription is generally considered as very inefficient. This, and the expected lability of RNA molecules, most probably account for the fact that in studies where sequencing of individual viruses is performed, viral RNA is first bulk transcribed to cDNA before proceeding to individual strand amplification. This strategy holds a risk for errors and *in vitro* recombination (Salazar-Gonzalez et al., 2008; Seifert et al., 2016), a risk that increases with transcript length. Fang et al. (1999) showed a 2.5-fold higher *in vitro* recombination rate for RT-PCR compared to PCR alone (6.49% *vs*. 2.65%) when amplifying a fragment of 4.5 kb. Di Giallonardo et al. (2013) reported an overall *in vitro* recombination frequency after RT-PCR of 53.6% when using a mixture of HIV-1 RNA, though a contribution of the RT step on the *in vitro* recombination was not demonstrated, probably because of the very short amplicon length (290 bp). In other reports, the estimated *in vitro* recombination frequencies during PCR ranged from < 1% to 7% (Bradley and Hillis, 1997; Judo et al., 1998; Meyerhans et al., 1990; Mild et al., 2011; Yang et al., 1996) but frequencies as high as 30% have been reported (Cronn et al., 2002; Yu et al., 2006).

We constructed a plasma pool with equal concentrations of subtype B, C and CRF02_AG virus and explored the possibility to dilute the

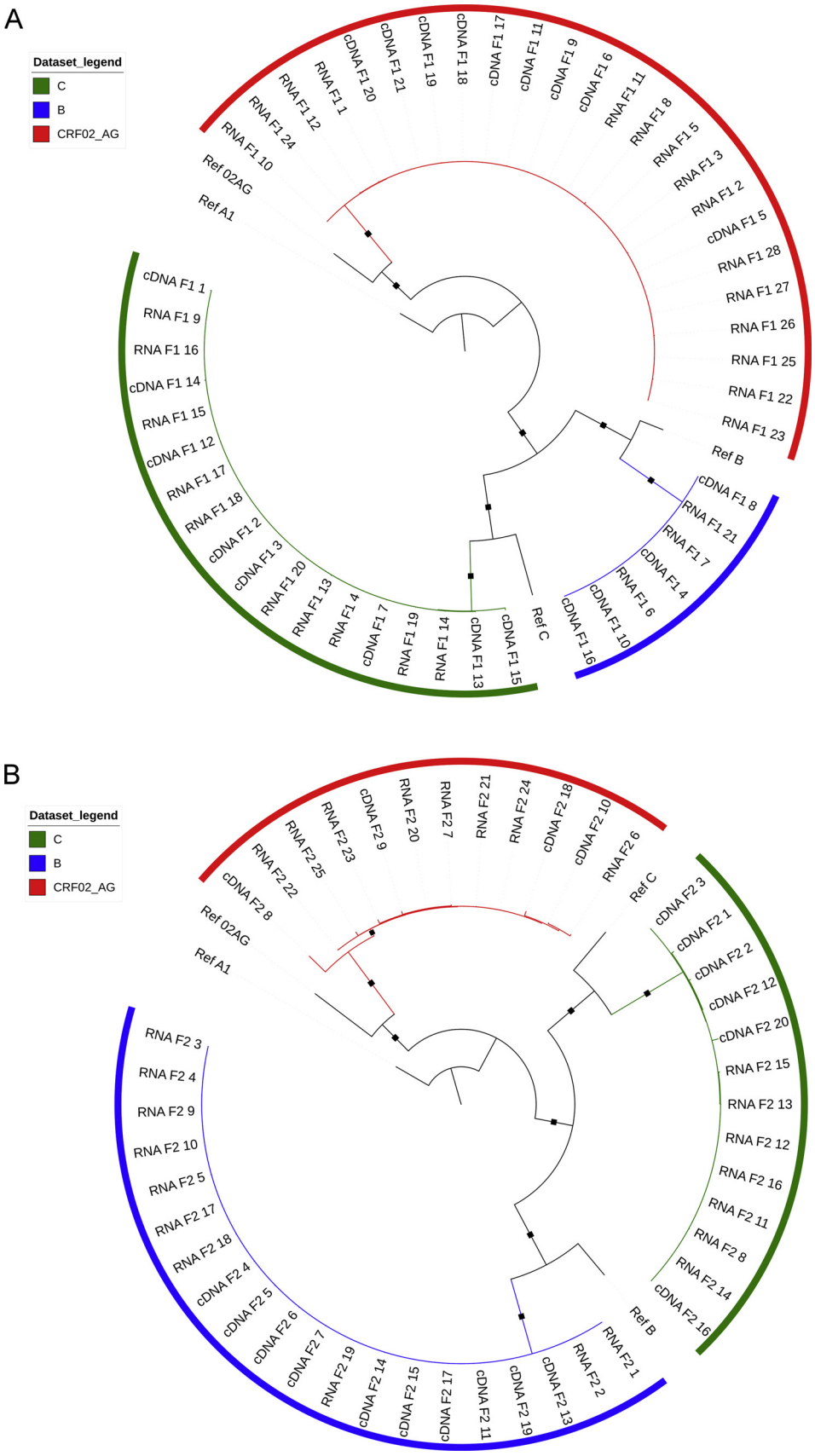

Fig. 2. Phylogenetic trees of sequences generated after limiting dilution of HIV-1 RNA or cDNA from a plasma pool with 3 HIV-1 variants. Phylogenetic tree of the 5′ amplicon sequences (A) and of the 3′ amplicon sequences (B). The tree is rooted on a A1 reference sequence (Ref_A1). A reference sequence was also added for each of the 3 subtypes (Ref_B, Ref_C, Ref_02AG). Branches are colored according to the subtype (02_AG: red, B: blue, C: green). Sequences are identified with the target of limiting dilution (RNA or cDNA), the fragment (F1: 5′ amplicon; F2: 3′ amplicon) and a serial number. aLTR SH-like node support values > 0.98 are indicated with black squares.

extracted viral RNA before cDNA generation in an attempt to prevent *in vitro* recombination. The results evidenced the feasibility of this strategy. *In vitro* recombination could be documented in one amplicon generated after cDNA dilution but no evidence of *in vitro* recombination was found in the amplicons obtained after RNA limiting dilution. Gall et al. (2012) used endpoint dilution of cDNA from a mixture of HIV-1 subtype B and C to assess occurrence of *in vitro* recombination. They detected 1 recombinant among 64 individual sequences (1.56%). It is clear to say that both in the study of Gall et al. and in our study, *in vitro* recombination remained rare but occurred and thus definitely constitutes a potential bias in studies on viral diversity and evolution (Gall et al., 2012).

The rationale behind limiting dilution as a way to perform clonal amplification is that there is a calculated chance of 80% that a positive reaction derives from a single molecule when a maximum of one-third of replicate reactions is positive (Rodrigo et al., 1997). This however assumes that the amplification reaction has a one copy sensitivity. It is unlikely that this is the case in our experiments, considering the very long amplification length. Based on the initially measured viral load of the plasma pool and assuming that no RNA is lost during the extraction process, each replicate endpoint dilution reaction contained a calculated equivalent of 32.8 RNA copies. More than one target RNA molecule may thus be present in a single reaction tube after limiting dilutions and *in vitro* recombination between these molecules cannot be fully excluded, though is expected to be rare. Lower target RNA concentrations have been associated with a decreased artificial chimera formation in RT-PCR (Waugh et al., 2015). We used Superscript III, a reverse transcriptase shown to be very sensitive and reliable for low copy number RNA samples (Okello et al., 2010). In the latter study, Superscript III consistently allowed amplification of 10 copies of an armored RNA preparation but the length of the amplified fragment (142 nt) was considerably smaller than the 4500 nt in our experiments. Using Superscript III, Palmer et al. (2003) succeeded in reverse transcribing and amplifying single copies of RNA transcripts from a plasmid containing a HIV-1 *gag* fragment of 554 bp.

We have no idea about the amount of RNA that is lost or degenerated during the extraction procedure. We used a manual extraction with the QIAamp Viral RNA Mini kit (Qiagen). In a recently published comparison of HIV-1 RNA extraction protocols for subsequent high-throughput sequencing, this kit produced the highest number of successfully amplified long amplicons (Cornelissen et al., 2017).

An important additional prerequisite for procedures that aim at characterizing complex virus populations is that the final result adequately reflects the variant composition in the initial sample. Whether using limiting dilution of RNA or limiting dilution of cDNA, all three subtypes present in the sample pool were recovered. Some differences in the overall distribution of the subtypes were noticed but these were most probably due to stochastic variation considering the low number of individual variants analyzed. There were no indications for a systematic over or underreporting of a particular subtype. If the results for all amplicons were added, the distribution was 35 (37.2%) for the CRF02_AG strain, 31 (33.0%) for subtype C and 27 (28.7%) for subtype B. Gall et al. (2012) sequenced 64 amplified products of a 1:1 plasma mixture (subtype B and C) using endpoint dilution and observed a distribution of 71.9% and 26.6% in relation to the initial viral loads of respectively 360,000 c/ml and 120,000 c/ml. It is clear that the accuracy with which the initial variability is grabbed will increase with increasing numbers of individual variants analyzed. In general, however typically 20–40 single genomes are analyzed in limiting dilution PCR experiments (Palmer et al., 2003). Possible founder effects can be limited by pooling a number of independent RNA isolations before proceeding to the RNA dilution as we did for the plasma pool.

For this study, we mainly concentrated on the generation of amplicons and less on the optimization of the sequencing reaction. For the sequencing reaction, 18 primers were used per amplicon. In ideal circumstances, this allowed to cover both amplicons bi-directionally. Problems were however encountered for the *envelope* region for some samples. To overcome gaps in the *envelope* sequence, alternative sequencing primers were designed.

An important limitation of the protocol developed is that the overlap between the two amplicons is short (105 nt) and does not allow the reliable linkage of the sequence obtained for both fragments to a single RNA molecule. Also, in order to achieve individual variant amplification we used a nested amplification protocol which may increase the risk of contamination. Careful examination of phylogenetic trees constructed with all obtained sequences could not show any evidence of contamination in our experiments.

In summary, this study presents a method for individual variant sequencing of the near full-length HIV-1 genome using two amplicons. The method is able to generate individual sequences from plasma samples with a viral load of less than 1000 c/ml. The method is unique in that it allows to link genomic regions over a distance covering half of the viral genome. The possibility to dilute the viral RNA to the extreme before reverse transcription was successfully explored as a way to sequence individual variants while limiting the chance for *in vitro* recombination. The protocol was set up using Sanger sequencing but adaptation of the limiting dilution PCR for preparation of NGS libraries must be feasible.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The AIDS Reference Laboratory of Ghent University is supported by the Belgian Ministry of Social Affairs through a fund within the Health Insurance System.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jviromet.2019.113737.

References

Alampalli, S.V., Thomson, M.M., Sampathkumar, R., Sivaraman, K., U, K.J.A, Dhar, C., G, D.S, Berry, N., Vyakarnam, A., 2017. Deep sequencing of near full-length HIV-1 genomes from plasma identifies circulating subtype C and infrequent occurrence of AC recombinant form in Southern India. PLoS One 12, e0188603.

Alizon, M., Montagnier, L., 1987. Genetic variability in human immunodeficiency viruses. Ann. N. Y. Acad. Sci. 511, 376–384.

Aralaguppe, S.G., Siddik, A.B., Manickam, A., Ambikan, A.T., Kumar, M.M., Fernandes, S.J., Amogne, W., Bangaruswamy, D.K., Hanna, L.E., Sonnerborg, A., Neogi, U., 2016. Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. J. Virol. Methods 236, 98–104.

Berg, M.G., Yamaguchi, J., Alessandri-Gradt, E., Tell, R.W., Plantier, J.C., Brennan, C.A., 2016. A Pan-HIV strategy for complete genome sequencing. J. Clin. Microbiol. 54, 868–882.

Bimber, B.N., Dudley, D.M., Lauck, M., Becker, E.A., Chin, E.N., Lank, S.M., Grunenwald, H.L., Caruccio, N.C., Maffitt, M., Wilson, N.A., Reed, J.S., Sosman, J.M., Tarosso, L.F., Sanabani, S., Kallas, E.G., Hughes, A.L., O'Connor, D.H., 2010. Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. J. Virol. 84, 12087–12092.

Bradley, R.D., Hillis, D.M., 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14, 592–593.

Brener, J., Gall, A., Batorsky, R., Riddell, L., Buus, S., Leitman, E., Kellam, P., Allen, T., Goulder, P., Matthews, P.C., 2015. Disease progression despite protective HLA expression in an HIV-infected transmission pair. Retrovirology 12, 55.

Cornelissen, M., Gall, A., Vink, M., Zorgdrager, F., Binter, S., Edwards, S., Jurriaans, S., Bakker, M., Ong, S.H., Gras, L., van Sighem, A., Bezemer, D., de Wolf, F., Reiss, P., Kellam, P., Berkhout, B., Fraser, C., van der Kuyl, A.C., Consortium, B., 2017. From clinical sample to complete genome: comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing. Virus Res. 239, 10–16.

Cromer, D., Grimm, A.J., Schlub, T.E., Mak, J., Davenport, M.P., 2016. Estimating the in-

vivo HIV template switching and recombination rate. AIDS 30, 185–192.
Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104, 482–489.
Cuevas, J.M., Geller, R., Garijo, R., Lopez-Aldeguer, J., Sanjuan, R., 2015. Extremely high mutation rate of HIV-1 in vivo. PLoS Biol. 13, e1002251.
Di Giallonardo, F., Zagordi, O., Duport, Y., Leemann, C., Joos, B., Kunzli-Gontarczyk, M., Bruggmann, R., Beerenwinkel, N., Gunthard, H.F., Metzner, K.J., 2013. Next-generation sequencing of HIV-1 RNA genomes: determination of error rates and minimizing artificial recombination. PLoS One 8, e74249.
Dilernia, D.A., Chien, J.T., Monaco, D.C., Brown, M.P., Ende, Z., Deymier, M.J., Yue, L., Paxinos, E.E., Allen, S., Tirado-Ramos, A., Hunter, E., 2015. Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. Nucleic Acids Res. 43, e129.
Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., et al., 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance–a review. Gene 40, 1–8.
Dukhovlinova, E., Masharsky, A., Vasileva, A., Porrello, A., Zhou, S., Toussova, O., Verevochkin, S., Akulova, E., Frishman, D., Montefiori, D., Labranche, C., Hoffman, I., Miller, W., Cohen, M.S., Kozlov, A.P., Swanstrom, R., 2018. Characterization of the transmitted virus in an ongoing HIV-1 epidemic driven by injecting drug use. AIDS Res. Hum. Retroviruses 34, 867–878.
Fang, G., Weiser, B., Visosky, A., Moran, T., Burger, H., 1999. PCR-mediated recombination: a general method applied to construct chimeric infectious molecular clones of plasma-derived HIV–1 RNA. Nat. Med. 5, 239.
Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., Berry, N., Pillay, D., Kellam, P., 2012. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. J. Clin. Microbiol. 50, 3838–3844.
Gorzer, I., Guelly, C., Trajanoski, S., Puchhammer-Stockl, E., 2010. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. J. Virol. Methods 169, 248–252.
Grossmann, S., Nowak, P., Neogi, U., 2015. Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management. J. Int. AIDS Soc. 18, 20035.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Judo, M.S., Wedel, A.B., Wilson, C., 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26, 1819–1825.
Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–5.
Liu, J., Song, H., Liu, D., Zuo, T., Lu, F., Zhuang, H., Gao, F., 2014. Extensive recombination due to heteroduplexes generates large amounts of artificial gene fragments during PCR. PLoS One 9, e106658.
Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.
Meyerhans, A., Vartanian, J.P., Wain-Hobson, S., 1990. DNA recombination during PCR. Nucleic Acids Res. 18, 1687–1691.
Mild, M., Hedskog, C., Jernberg, J., Albert, J., 2011. Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. PLoS One 6, e22741.
Miranda, J.A., Steward, G.F., 2017. Variables influencing the efficiency and interpretation of reverse transcription quantitative PCR (RT-qPCR): an empirical study using Bacteriophage MS2. J. Virol. Methods 241, 1–10.
Nadai, Y., Eyzaguirre, L.M., Constantine, N.T., Sill, A.M., Cleghorn, F., Blattner, W.A., Carr, J.K., 2008. Protocol for nearly full-length sequencing of HIV-1 RNA from plasma. PLoS One 3, e1420.
Ode, H., Matsuda, M., Matsuoka, K., Hachiya, A., Hattori, J., Kito, Y., Yokomaku, Y., Iwatani, Y., Sugiura, W., 2015. Quasispecies analyses of the HIV-1 near-full-length genome with illumina MiSeq. Front. Microbiol. 6, 1258.
Okello, J.B., Rodriguez, L., Poinar, D., Bos, K., Okwi, A.L., Bimenya, G.S., Sewankambo, N.K., Henry, K.R., Kuch, M., Poinar, H.N., 2010. Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. PLoS One 5, e13931.
Palmer, S., Wiegand, A.P., Maldarelli, F., Bazmi, H., Mican, J.M., Polis, M., Dewar, R.L., Planta, A., Liu, S., Metcalf, J.A., Mellors, J.W., Coffin, J.M., 2003. New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J. Clin. Microbiol. 41, 4531–4536.
Posada-Cespedes, S., Seifert, D., Beerenwinkel, N., 2017. Recent advances in inferring viral diversity from high-throughput sequencing data. Virus Res. 239, 17–32.
Rodrigo, A.G., Goracke, P.C., Rowhanian, K., Mullins, J.I., 1997. Quantitation of target molecules from polymerase chain reaction-based limiting dilution assays. AIDS Res. Hum. Retroviruses 13, 737–742.
Rousseau, C.M., Birditt, B.A., McKay, A.R., Stoddard, J.N., Lee, T.C., McLaughlin, S., Moore, S.W., Shindo, N., Learn, G.H., Korber, B.T., Brander, C., Goulder, P.J., Kiepiela, P., Walker, B.D., Mullins, J.I., 2006. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. J. Virol. Methods 136, 118–125.
Salazar-Gonzalez, J.F., Bailes, E., Pham, K.T., Salazar, M.G., Guffey, M.B., Keele, B.F., Derdeyn, C.A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J.A., Swanstrom, R., Haynes, B.F., Athreya, G.S., Korber, B.T., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J. Virol. 82, 3952–3970.
Salazar-Gonzalez, J.F., Salazar, M.G., Keele, B.F., Learn, G.H., Giorgi, E.E., Li, H., Decker, J.M., Wang, S., Baalwa, J., Kraus, M.H., Parrish, N.F., Shaw, K.S., Guffey, M.B., Bar, K.J., Davis, K.L., Ochsenbauer-Jambor, C., Kappes, J.C., Saag, M.S., Cohen, M.S., Mulenga, J., Derdeyn, C.A., Allen, S., Hunter, E., Markowitz, M., Hraber, P., Perelson, A.S., Bhattacharya, T., Haynes, B.F., Korber, B.T., Hahn, B.H., Shaw, G.M., 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J. Exp. Med. 206, 1273–1289.
Schultz, A.K., Bulla, I., Abdou-Chekaraou, M., Gordien, E., Morgenstern, B., Zoaulim, F., Deny, P., Stanke, M., 2012. jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus. Nucleic Acids Res. 40, W193–W198.
Seifert, D., Di Giallonardo, F., Topfer, A., Singer, J., Schmutz, S., Gunthard, H.F., Beerenwinkel, N., Metzner, K.J., 2016. A comprehensive analysis of primer IDs to study heterogeneous HIV-1 populations. J. Mol. Biol. 428, 238–250.
Struck, D., Lawyer, G., Ternes, A.M., Schmit, J.C., Bercoff, D.P., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42, e144.
Svarovskaia, E.S., Cheslock, S.R., Zhang, W.H., Hu, W.S., Pathak, V.K., 2003. Retroviral mutation rates and reverse transcriptase fidelity. Front. Biosci. 8, d117–d134.
Tovanabutra, S., Sanders, E.J., Graham, S.M., Mwangome, M., Peshu, N., McClelland, R.S., Muhaari, A., Crossler, J., Price, M.A., Gilmour, J., Michael, N.L., McCutchan, F.M., 2010. Evaluation of HIV type 1 strains in men having sex with men and in female sex workers in Mombasa, Kenya. Aids Res. Hum. Retrov. 26, 123–131.
Verhofstede, C., Fransen, K., Van Den Heuvel, A., Van Laethem, K., Ruelle, J., Vancutsem, E., Stoffels, K., Van den Wijngaert, S., Delforge, M.L., Vaira, D., Hebberecht, L., Schauvliege, M., Mortier, V., Dauwe, K., Callens, S., 2017. Decision tree for accurate infection timing in individuals newly diagnosed with HIV-1 infection. BMC Infect. Dis. 17, 738.
Walker, C.M., Henn, M.R., Boutwell, C.L., Charlebois, P., Lennon, N.J., Power, K.A., Macalalad, A.R., Berlin, A.M., Malboeuf, C.M., Ryan, E.M., Gnerre, S., Zody, M.C., Erlich, R.L., Green, L.M., Berical, A., Wang, Y., Casali, M., Streeck, H., Bloom, A.K., Dudek, T., Tully, D., Newman, R., Axten, K.L., Gladden, A.D., Battis, L., Kemper, M., Zeng, Q., Shea, T.P., Gujja, S., Zedlack, C., Gasser, O., Brander, C., Hess, C., Günthard, H.F., Brumme, Z.L., Brumme, C.J., Bazner, S., Rychert, J., Tinsley, J.P., Mayer, K.H., Rosenberg, E., Pereyra, F., Levin, J.Z., Young, S.K., Jessen, H., Altfeld, M., Birren, B.W., Walker, B.D., Allen, T.M., 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.
Waugh, C., Cromer, D., Grimm, A., Chopra, A., Mallal, S., Davenport, M., Mak, J., 2015. A general method to eliminate laboratory induced recombinants during massive, parallel sequencing of cDNA library. Virol. J. 12, 55.
Yang, Y.L., Wang, G., Dorman, K., Kaplan, A.H., 1996. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. AIDS Res. Hum. Retroviruses 12, 303–306.
Yu, W., Rusterholtz, K.J., Krummel, A.T., Lehman, N., 2006. Detection of high levels of recombination generated during PCR amplification of RNA templates. Biotechniques 40, 499–507.
Zagordi, O., Klein, R., Daumer, M., Beerenwinkel, N., 2010. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res. 38, 7400–7409.
Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J., Neher, R.A., 2015. Population genomics of intrapatient HIV-1 evolution. Elife 4.