From clinical sample to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing

Marion Cornelissen [a,1], Astrid Gall [b,1], Monique Vink [a], Fokla Zorgdrager [a], Špela Binter [b], Stephanie Edwards [b], Suzanne Jurriaans [c], Margreet Bakker [a], Swee Hoe Ong [b], Luuk Gras [d], Ard van Sighem [d], Daniela Bezemer [d], Frank de Wolf [e], Peter Reiss [d,f,g], Paul Kellam [b], Ben Berkhout [a], Christophe Fraser [e], Antoinette C. van der Kuyl [a,*], the BEEHIVE Consortium

[a] Laboratory of Experimental Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center of the University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands
[b] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom
[c] Laboratory of Clinical Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center of the University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands
[d] Stichting HIV Monitoring, 1105 BD Amsterdam, The Netherlands
[e] Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London W21PG, UK
[f] Department of Global Health, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands
[g] Amsterdam Institute of Global Health and Development, 1105 BM Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

The BEEHIVE (Bridging the Evolution and Epidemiology of HIV in Europe) project aims to analyse nearly-complete viral genomes from >3000 HIV-1 infected Europeans using high-throughput deep sequencing techniques to investigate the virus genetic contribution to virulence. Following the development of a computational pipeline, including a new de novo assembler for RNA virus genomes, to generate larger contiguous sequences (contigs) from the abundance of short sequence reads that characterise the data, another area that determines genome sequencing success is the quality and quantity of the input RNA. A pilot experiment with 125 patient plasma samples was performed to investigate the optimal method for isolation of HIV-1 viral RNA for long amplicon genome sequencing. Manual isolation with the QIAamp Viral RNA Mini Kit (Qiagen) was superior over robotically extracted RNA using either the QIAcube robotic system, the *m*Sample Preparation Systems RNA kit with automated extraction by the m2000*sp* system (Abbott Molecular), or the MagNA Pure 96 System in combination with the MagNA Pure 96 Instrument (Roche Diagnostics). We scored amplification of a set of four HIV-1 amplicons of ~1.9, 3.6, 3.0 and 3.5 kb, and subsequent recovery of near-complete viral genomes.

Subsequently, 616 BEEHIVE patient samples were analysed to determine factors that influence successful amplification of the genome in four overlapping amplicons using the QIAamp Viral RNA Kit for viral RNA isolation. Both low plasma viral load and high sample age (stored before 1999) negatively influenced the amplification of viral amplicons >3 kb. A plasma viral load of >100,000 copies/ml resulted in successful amplification of all four amplicons for 86% of the samples, this value dropped to only 46% for samples with viral loads of <20,000 copies/ml.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Enabled by the recent developments in high-throughput sequencing techniques, the molecular analysis of complete or nearly-complete viral genomes, including HIV, is now becoming the new research standard. Complete genomes contain more information on e.g. viral virulence elements than the shorter

genome fragments that were previously investigated using Sanger dideoxynucleotide chain terminator sequencing methods. However, assembly of complete viral genomes from the relatively short reads generated by most high-throughput sequencing systems has proven to be a challenge, and much effort has been directed towards optimization of this part of the process. The computational tools to process high-throughput sequencing data are now coming of age, so that it is now time to critically examine earlier steps in the process. Many experimental steps have to be performed between the isolation of nucleic acids from a patient sample and the investigation of the assembled HIV-1 genomes. Many of these are potentially executed under suboptimal conditions that need to be improved. Several factors have to be taken into account, including sample storage, storage temperature, storage time, number of freeze-thaw cycles, RNA/DNA extraction methods, and selection of the primers as well as the enzymes for reverse transcription and PCR amplification. For RNA viruses, selection of the reverse-transcriptase (RT) enzyme and primers for the RT-reaction are major concerns. For genetically variable viruses such as Human Immunodeficiency Virus type 1 (HIV-1) and Hepatitis C Virus (HCV), the primers used for subsequent DNA amplification are also important. Within an infected patient, diversification of the virus population during chronic infection increases the likelihood that primers designed to amplify early virus variants may not work optimally on subsequent variants. In addition, the plasma viral load that primarily determines the amount of starting material for the RT reaction can differ significantly between patients. If the input material for high-throughput sequencing, the amplified cDNA/DNA, is of low yield and/or poor quality, it follows that investing time and money into optimising assembly of poor quality data is not cost effective and almost impossible in high-throughput settings.

The BEEHIVE (Bridging the Evolution and Epidemiology of HIV in Europe) project is a large scale study to determine the viral genetic basis of HIV-1 virulence using genome high-throughput sequencing. The project aims to collect nearly-complete viral genomes from >3000 HIV-1 infected European individuals who are either recent seroconverters or have presented with an acute HIV infection, enabling an estimate of the duration of infection. Samples are collected from participating hospitals and institutes in six European countries. For each eligible patient, the first set-point viral load sample has been retrieved. Viral RNA is extracted at a central laboratory (AMC, Amsterdam, The Netherlands) and subsequently sequenced at the Wellcome Trust Sanger Institute (Hinxton, UK), where the RNA is further processed. To investigate the effect of different nucleic acid isolation methods, and manual versus robotic extraction, 125 samples were first examined in five pilot experiments. After identification of the optimal isolation technique, 616 further BEEHIVE samples were analysed. Amplification and sequencing results were compared to sample characteristics such as age and the number of viral copies used as input material for the RT reaction. In addition, an overview of nucleic acid isolation methods used in literature for HIV-1 complete genome high-throughput sequencing will be provided.

2. Materials and methods

2.1. Patient material

Patients eligible for inclusion in the BEEHIVE study have to meet the following criteria: (i) a known seroconversion interval of maximum one year between the last negative and the first positive HIV test or clear evidence of acute illness or recent HIV-1 infection at the first positive test. (ii) Patients should be anti-retroviral therapy (ART) naïve, i.e. have not taken therapy for the first six months following the first positive test. (iii) Patients should have

at least one viral load determination, or a sample that can be used for this purpose should be available, dating to 6–24 months after the first positive test and before the start of any ART. (iv) Lastly, a sample of at least 500 µl of frozen EDTA blood plasma or serum taken between 0 and 24 months following the first positive test should be available, also before the start of any ART. Samples were stored at $-80\,^{\circ}$C. The first samples included in BEEHIVE were from the Netherlands, where Stichting HIV Monitoring (SHM; HIV Monitoring Foundation) acts as the national reference centre collecting data from all HIV-positive individuals in care. In the Netherlands, most HIV-1 positive samples meeting the above criteria are from men who have sex with men (MSM) that are mostly infected with HIV-1 subtype B; in the 1980s 100% of MSM were infected with subtype B, in November 2011 this figure was reduced to 77% subtype B infections (Bezemer et al., 2015; van der Kuyl et al., 2013). All HIV-1 RNA isolations for the BEEHIVE study were performed at the Department of Medical Microbiology of the Academic Medical Center in Amsterdam, the Netherlands.

2.2. HIV-1 RNA isolation methods

Viral RNA isolation methods tested in the pilot experiments were the QIAamp Viral RNA Mini Kit that uses spin columns to purify the RNA (Qiagen, Venlo, the Netherlands), the *m*Sample Preparation Systems RNA kit for sample preparation with automated extraction by the m2000*sp* system (Abbott Molecular, Des Plaines, IL, USA), and the MagNA Pure 96 System in combination with the MagNA Pure 96 Instrument (Roche Diagnostics Nederland, Almere, Netherlands). The QIAamp Viral RNA isolation was done either manually or using the QIAcube robotic system, which is designed for fully automated sample preparation with the QIAamp RNA isolation kits. All RNA extractions were performed according to the manufacturers' instructions. For details of the isolation methods, see Supplementary Table 1. As input material, 200–250 µl of blood plasma or serum was used, regardless of the viral genome copy number; for the automated extractions, the input volume was 1 ml.

2.3. Quality control of the isolated HIV-1 RNA

To assess the quality of the isolated viral RNA, 10 µl (1/8) was used in an RT-PCR using the SuperscriptIII One-step RT-PCR System with Platinum *Taq* DNA polymerase (Invitrogen) performed according to the manufacturer's instructions. Primers used were 5′ED31 5′-CCTCAGCCATTACACAGGCCTGTCCAAAG -3′ (Delwart et al., 1995) and 3′A1191 (5′-AGCAATGTATGCCCCTCCCAT-3′, position 7510-7531 of the HXB2 reference strain with GenBank accession number K03455) which target the HIV-1 *envelope* gene to generate a product of 725 base pairs (bp). PCR products were analysed on a 2% agarose gel. The HIV-1 *envelope* product could be amplified from 98.5% of the RNA samples. Samples that failed this quality control were not used for subsequent complete genome RT-PCR amplification and are not included in the numbers reported in this study.

2.4. RT-PCR amplification and high-throughput sequencing of nearly-complete HIV-1 genomes

Four 5 µl aliquots of viral RNA were reverse transcribed and amplified using the SuperScriptIII One-Step RT-PCR system with Platinum *Taq* DNA High Fidelity polymerase (Invitrogen) using four different primer sets (Gall et al., 2012). The *pan*-HIV-1 specific primer sets target semi-conserved regions of the genome and were developed using an alignment of approximately 1500 HIV-1 genome sequences (Gall et al., 2012). The primer sets amplify four overlapping amplicons that span the entire protein-coding

ARTICLE IN PRESS

Table 1

Outcomes of the pilot experiment and five initial BEEHIVE sample sets.

Sample batch	No.	Viral loads (copies/ml)	Year of isolation	RNA extraction	RT-PCR results (no. of amplicons)						Sequencing method	Complete HIV-1 genomes	
					0–2		3		4				
					No.	%	No.	%	No.	%		No.	%
Pilot 1	45	12,572–2,800,000	1996–2010	QIAamp Viral RNA [manual]	0	0	5	11.1	40	88.9	Roche/454	38	84.4
Pilot 2	24	4037–1,079,658	2012	M2000 (Abbott)	24	100	0	0	0	0	N/A	N/A	N/A
Pilot 3a[a]	12	14,100–1,929,719	2012	M2000 (Abbott)	11	91.7	1	8.3	0	0	N/A	N/A	N/A
Pilot 3b[a]	12	14,100–1,929,719	2012	MagNa Pure (Roche)	9	75	1	8.3	2	16.7	N/A	N/A	N/A
Pilot 4	20	19,508–1,012,000	2012	MagNa Pure (Roche)	20	100	0	0	0	0	N/A	N/A	N/A
Pilot 5	12	13,514–173,219	2012	QIAcube + QIAamp Viral RNA	4	33.3	4	33.3	4	33.3	MiSeq 150 bp PE	4	33.3
Total	125												
BEEHIVE 1	48	7566–1,744,469	1992–2009	QIAamp Viral RNA [manual]	4	8.3	9	18.8	35	72.9	MiSeq 150 bp PE	35	72.9
BEEHIVE 2[b]	103	4200–978,120	1999–2011	QIAamp Viral RNA [manual]	3	2.9	13	12.6	87	84.5	MiSeq 250 bp PE	76	73.8
BEEHIVE 3[b]	206	1160–1,012,000	1985–2011	QIAamp Viral RNA [manual]	59	28.6	49	23.8	98	47.6	MiSeq 300 bp PE	88	42.7
BEEHIVE 4[b]	159	40–2,730,000	1985–2011[c]	QIAamp Viral RNA [manual]	2	1.3	15	9.4	142	89.3	MiSeq 300 bp PE	132	83
BEEHIVE 5[b]	100	40–3,681,880	1998–2011	QIAamp Viral RNA [manual]	9	9	9	9	82	82	MiSeq 300 bp PE	73	73
Total	616				77	12.5	95	15.4	444	72.1		404	65.6

N/A = not applicable.

[a] The same sample set was used in pilot 3a and 3b.

[b] Pre-screened for RNA quality by env RT-PCR.

[c] One sample was from 1985, the remaining samples were from 1999 or later.

4 M. Cornelissen et al. / Virus Research xxx (2016) xxx–xxx

sequence and almost the complete HIV-1 RNA genome. Amplicon pan-HIV-1_1 (5′LTR-*gag*) is approximately 1.9 kb, amplicon pan-HIV-1_2 (*gag-pol*) is 3.6 kb, amplicon pan-HIV-1_3 (*pol-env*) is 3 kb, and amplicon pan-HIV-1_4 (*tat-*3′LTR) is 3.5 kb. For Roche/454 sequencing, amplicons were pooled in equimolar amounts. Single-stranded DNA libraries were prepared from 500 ng DNA with the GS FLX Titanium Rapid Library Preparation Kit according to the manufacturer's instructions, using one of the 12 Multiplex Identifier (MID) adaptors for each sample. Sequencing was performed using the Genome Sequencer FLX Instrument and GS FLX Titanium series reagents as described previously (Gall et al., 2012). For Illumina sequencing, 5 μl of amplicon I were pooled with 10 μl each of amplicons II–IV. Libraries were prepared from 50 to 1000 ng DNA as described in Quail et al. (2008, 2012), using one of 96 multiplex adaptors for each sample. Paired-end sequencing was performed using the Illumina MiSeq instrument as described previously (Gall et al., 2014), with read lengths as outlined in Table 1.

2.5. Virus genome assembly and determination of the HIV-1 subtype

Assembly of HIV-1 genomes from Roche/454 data was performed as described previously (Gall et al., 2012), while the de novo assembler IVA (Iterative Virus Assembler, (Hunt et al., 2015)) was used for Illumina MiSeq data. A genome assembly was defined as "nearly-complete" if (1) the contigs cover at least 96% of the amplifiable genome, demarcated by the two outermost RT-PCR primers, and (2) read coverage was present before or at the start of the *gag* gene and reached the end of the *nef* gene. Subtyping was done using COMET at the contig level (Struck et al., 2014). Samples resulting in a single contig classified as "unassigned" by COMET, and samples resulting in multiple contigs classified as different subtypes by COMET, were categorised as "unassigned".

2.6. Overview of HIV-1 RNA isolation methods used in the literature

A survey of HIV-1 RNA isolation methods for high-throughput sequencing was performed using the PubMed database at www.ncbi.nlm.nih.gov/pubmed with the search terms HIV-1; RNA isolation; complete genome; clinical samples; next-generation sequencing, high-throughput sequencing; Illumina MiSeq; Roche 454; PacBio; Ion Torrent, in various combinations.

3. Results

3.1. Comparison of viral RNA isolation techniques

To investigate the optimal RNA isolation protocol for high-throughput HIV-1 sequencing, a total of 125 plasma or serum samples were processed in five pilot experiments. The number of clinical samples used in each pilot ranged from 12 to 45, and the viral loads varied from low (±12,000 copies/ml) to high (1–3 × 10⁶ copies/ml). Three automated extraction systems with corresponding isolation kits from different suppliers were used; all reagents and machines are commercially available. In addition, manual extractions were performed with the QIAamp Viral RNA Mini Kit. The results of the pilot experiments are summarized in Table 1. To be able to reconstruct near-complete HIV-1 genomes, it is necessary that all four RT-PCR amplicons are amplified. The number of successfully amplified amplicons from manually purified RNA isolations was greater than that from all robotically isolated RNA preparations. For RNA isolated with the m2000*sp* and MagNa Pure systems, more than two RT-PCR amplicons could be generated only rarely, precluding complete HIV-1 genome sequencing. The best robotic RNA isolation method used the QIAamp Viral RNA Kit in
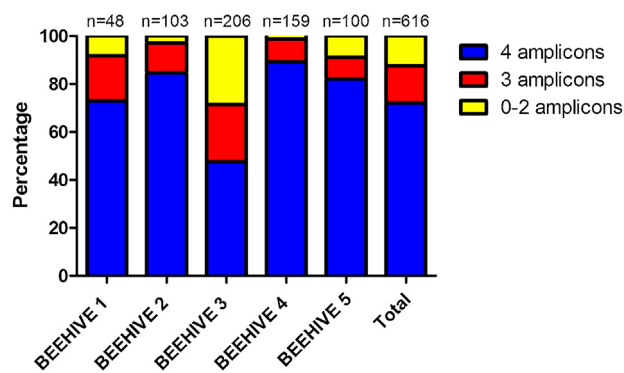

Fig. 1. RT-PCR success rate by BEEHIVE sample subset. Percentages of samples that generated 0–2, 3, or 4 HIV-1 amplicons are presented for the five BEEHIVE pilot experiments.
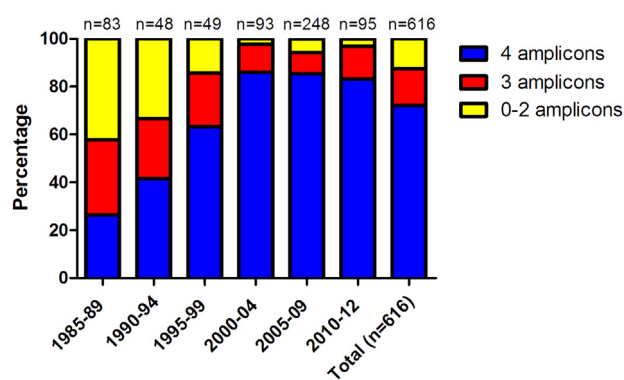

Fig. 2. RT-PCR success rate and sample age. The percentage of BEEHIVE samples that generated 0–2, 3, or 4 HIV-1 amplicons is provided per five-year sampling period (three year period for the most recent category from 2010 to 2012), ranging from 1985 to 2012.

combination with the automated QIAcube system. Based on these results, it was decided to use manual extraction with the QIAamp Viral RNA Kit as the method of choice for the BEEHIVE study, notwithstanding the additional time and staff required.

3.2. Sample characteristics and HIV-1 complete genome sequencing

In five subsequent experiments, HIV-1 RNA was manually isolated, amplified by RT-PCR and sequenced from a total of 616 Dutch patient samples (Table 1). The robustness of the chosen isolation method could thus be assessed, and sample characteristics associated with RT-PCR performance were investigated. Most samples (76.5%) contained subtype B viruses. The overall RT-PCR success rate for the 616 samples, divided into five sets of 48–206 samples each, for generating the four amplicons ranged between 48 and 89% (mean 75%) (Fig. 1). The two subsets that performed below 80% in the amplifications contained 131 samples dating to the early years of the HIV-1 epidemic in the Netherlands, i.e. from 1985 to 1994 (Fig. 2). Older samples are more likely to contain degraded viral RNA due to storage conditions or multiple freeze-thaw cycles. In favour of RNA degradation is the observation that the smallest amplicon of 1.9 kb was amplified for the majority of samples irrespective of their age, while a lower success rate was scored for the three larger amplicons (result not shown).

ARTICLE IN PRESS

Table 2
Overview of HIV RNA isolation methods for full-length genome sequencing from clinical specimens.

Reference	Input patient material	RNA isolation method	(RT-)PCR strategy	Sequencing platform and concluding remarks
Gall et al., 2012	4 UK HIV-1 infected blood plasma samples. Viral load range 3000–300,000 copies/ml	Manual RNA isolation with the QIAamp Viral RNA Mini Kit	SuperScriptIII RT/ Platinum *Taq* High Fidelity. PCR reactions amplify four fragments	454 Life Sciences/Roche platform. All four fragments amplified, smallest fragment of 1,9 kb gives the best amplification. For all samples, the complete amino acid coding region of the HIV-1 genome was covered
Henn et al., 2012	89 Blood plasma samples of 83 HIV-1 subtype B infected participants of North American, Swiss and German cohorts	Manual RNA isolation with the QIAamp Viral RNA Mini Kit Pre-treatment: plasma samples (1 ml) were centrifuged for 1 h at 20.817 g.	Superscript III RT (Invitrogen) PCR reactions amplify four fragments of 3.2 kb each	454 Life Sciences/Roche platform. Only a few samples failed to amplify all four fragments with subsequent low sequence coverage
Giallonardo et al., 2014	2 Clinical samples (1 ml plasma) from a Swiss HIV-1 cohort	Manually with the NucleoSpin RNA Virus Kit (Macherey and Nagel, Oensingen) Pre-treatment: plasma samples (1 ml) were centrifuged for 1 h at 50.000 g.	PrimeScript Reverse Transcriptase (Clontech). Nested or semi-nested PCR amplifying 5 and 11 amplicons, respectively	Illumina MiSeq instrument as well as the Pacific Biosciences platform
Berg et al., 2016	8 HIV-1 infected blood plasma samples from blood banks in Cameroon, South Africa and Thailand. For the m2000*sp* extraction 500–600 μl of sample was used	RNA isolation with the RNA Sample Preparation System (Abbott Molecular; Des Plaines, IL) on a m2000*sp* instrument. Extractions were also performed using the MagMax (Life Technologies: Grand Island, NY), Viral Mini (Qiagen), Ultra Sens Virus (Qiagen), and PureLink (Life Technologies) kits with or without automated extractions	SMARTScribe Reverse Transcriptase (Clontech, Mountain View, CA). Subsequent amplification with a mixture containing 6 HIV specific primers	Illumina MiSeq instrument. All 8 clinical samples achieved between 95–100% genome coverage with average sequence depths ranging from 494× to 6481×
Brener et al., 2015	Transmission pair: 3 samples from donor and 3 from recipient. pVL range from 3000 to ±80,000 copies/ml.	Manual RNA isolation with the QIAamp Viral RNA Mini Kit. Pre-treatment: plasma samples (1 ml) were centrifuged for 1 h at 21.000 rpm. For samples with a pVL <3000 copies/ml 3 ml input volume was used	Superscript III One-Step RT-PCR with Platinum Taq High Fidelity Enzyme Mix (Invitrogen, Carlsbad, California, US).	Illumina MiSeq 250 bp paired end technology
Luk et al., 2015	35 Blood plasma samples. Specimens with high viral loads ranging from 4.45 to 5.90 log10 copies/mL were selected.	EZ1 Virus Mini Kit v2.0 protocol for Serum (Qiagen) with the EZ1 BioRobot robotic extraction	MMLV SuperScript III Reverse Transcriptase (Invitrogen/Life Technologies, Carlsbad, CA), using a primer that has a random 9-mer linked to a specific 17-mer sequence	Illumina MiSeq instrument. Greater than 90% genome coverage achieved for 31 (89%) of the 35 specimens, average sequence depth ranging from 37- to 8274-fold, permitting full length viral genome assembly for 25 (71%)
Ode et al., 2015	97 Blood plasma samples obtained from 58 Japanese patients were used. The plasma viral load ranged from 10–28,000 copies/ml	RNA extracted from 200 or 400 μl plasma with the MagNAPure Compact Nucleic Acid Isolation Kit I (Roche Diagnostics) and MagNA Pure Compact Instrument	RT reaction with PrimeScript II High Fidelity One Step RT-PCR Kit (Takara, Shiga, Japan). HIV-1 nucleotides 706-8825 (*gag-nef*) were amplified in four nested overlapping fragments	Illumina MiSeq instrument. The four segments were successfully amplified for the subtype B samples with >200 copies/ml. Of the 39 non-B subtype viruses, subtype F often failed in *int-env* fragment amplification
Zanini et al., 2015	9 Swedish HIV-1 infected patients, 7–12 blood plasma samples each, resulting in a total of 73 clinical samples. Samples dated to 1990–2003 and had always been stored at −70 °C. Plasma viral load range 0–648,000 copies/ml	Manual extraction with RNeasy Lipid Tissue Mini Kit (Qiagen)	Superscript III One-Step RT-PCR with Platinum Taq High Fidelity Enzyme Mix (Invitrogen, Carlsbad, California, US). Subsequent PCR amplification of six overlapping fragments, nt. 571-9567 of HXB2 reference strain	Illumina MiSeq instrument. One fragment had a significantly lower coverage than the other five. Lower viral loads generally resulted in less coverage

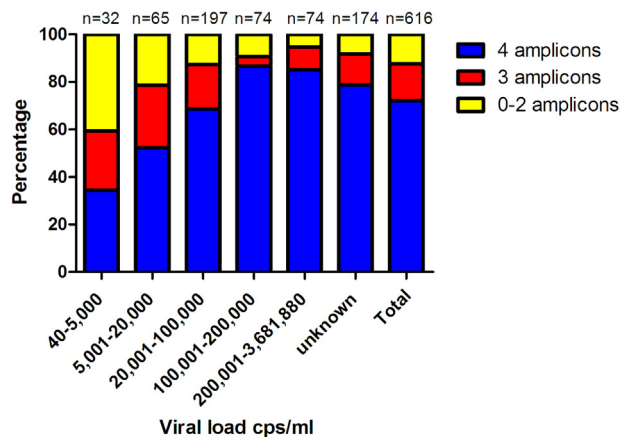## RT-PCR success rate and plasma viral load

Fig. 3. RT-PCR success rate and plasma viral load. The percentage of BEEHIVE samples that generated 0–2, 3, or 4 HIV-1 amplicons was calculated for five HIV-1 plasma viral load groups plus the samples with unknown viral load.

The initial viral load in the samples is another important factor for RT-PCR success rates, with lower viral loads expected to correlate with RT-PCR failure. Fig. 3 indicates that the success rate for amplifying all four amplicons was much lower for samples with lower viral loads ≤5000 copies/ml (34%) than for samples with high viral loads >100,000 copies/ml (86%). For samples with intermediate viral loads between 5000 and 100,000 copies/ml, the success rate was 65%.

Viral load was, similar to sample age, not a major factor in successful amplification of the shorter, 1.9 kb, amplicon, which could be amplified for 451/616 samples (73%). Individual success rates for the three larger amplicons were 65%, 62% and 64%, respectively.

3.3. HIV-1 RNA isolation methods used in other studies

A literature search for methods used to isolate HIV-1 RNA for complete genome high-throughput sequencing was performed. A total of eight publications were retrieved and inspected for the sample type involved, the RNA isolation method used and the outcomes of the sequencing protocol. The results are summarized in Table 2. All eight studies concerned blood plasma samples, none reported the use of serum. Numbers of samples analysed ranged from one to 97. Table 2 illustrates that the studies differed greatly in the reporting of details of the samples the actual results of sequencing. Only five publications provided the viral load of the samples, or more specifically, the input RNA copy numbers for the RT-PCR amplification. Two studies reported the use of machines for RNA isolation (Luk et al., 2015; Ode et al., 2015), one study apparently alternated between machine and manual extraction (Berg et al., 2016), while the remaining five studies exclusively used manual extraction (Brener et al., 2015; Gall et al., 2012; Giallonardo et al., 2014; Henn et al., 2012; Zanini et al., 2015).

4. Discussion

A total of 125 plasma samples from HIV-1-positive individuals, intended for high-throughput complete genome sequencing, were subjected to three different viral RNA isolation methods, the QIAamp Viral RNA Mini Kit (Qiagen), the *m*Sample Preparation Systems RNA Kit (Abbott Molecular), and the MagNA Pure 96 System (Roche Diagnostics). The latter two isolation kits are to be used with a machine developed by the supplier, while the QIAamp Viral RNA Mini Kit can either be used for manual RNA isolation or with a matching robotic system. Subsequently, isolated viral RNA was reverse-transcribed and used for the amplification of four overlapping RT-PCR amplicons that span almost the complete HIV-1 genome. Generation of all four amplicons is essential for successful sequencing of the near-complete viral genome. Isolating sufficient virus genome RNA molecules of sufficient length is an important factor in generating these RT-PCR amplicons, the second factor being the specific primer sets used. In earlier tests, the *pan*-HIV-1 primer sets employed here were used for successful amplification of HIV-1 group M, N, and O sequences, including all major group M subtypes and many circulating recombinant forms, either from cell culture reference viruses, primary clinical isolates and uncultured plasma virus samples (Gall et al., 2012). Therefore, we assumed that failure to amplify an HIV-1 amplicon in this pilot study is mainly due to insufficient quantity and/or quality of the isolated viral RNA, and less so due to non-matching RT-PCR primers. We conclude that generation of the complete set of four amplicons has the highest success rate when the QIAamp Viral RNA Kit is combined with manual RNA extraction. This isolation method also worked comparatively well with machine extraction, suggesting that the QIAamp Viral RNA Kit outperforms the isolation methods from other suppliers. The outcome may suggest that manual HIV-1 RNA isolation results in less shearing of the 9 kb genome than robotic extractions.

Next, we used the optimal isolation method for 616 plasma samples from HIV-1-positive patients included in the BEEHIVE study to assess what sample characteristics influence the reverse transcription and/or the amplification process. We report that both sample age and plasma viral load influence the RT-PCR success. The smallest amplicon of 1.9 kb was amplified for 73.2% of the samples, whereas the larger amplicons (>3 kb) exhibit reduced success rates between 61.9 and 65.1%. The observation that amplification results correlate with amplicon length suggests that differences in the number of intact RNA copies is more important than primer mismatches with the target sequence. A low input RNA concentration in the RT-PCR reaction is most likely due a low sample viral load or because of RNA degradation related to sample age, suboptimal storage and/or multiple freeze-thaw cycles.

Analysis of the literature indicated that a wealth of RNA isolation systems are used by others. As RT-PCR approaches differ and reporting of sequencing success was incomplete and not uniform, it is difficult to estimate the relative value of each RNA isolation method. Each study defines its own endpoint; for instance, the sequence coverage per position in the HIV-1 genome is not standardised as to what cut-off should be used for reliable assembly. Viral load is an important factor determining sequence success. However, for a number of samples with low viral loads sufficient sequence information was generated, while for some samples with high viral loads amplification steps failed, suggesting that a combination of RNA input and specific primers was more important than either of these factors alone. No study generated full-length sequences of the HIV-1 RNA genome. Due to the position of the RT-PCR primers, the complete long terminal repeats (LTR) are lacking in the study of Ode et al. (2015), while in the other seven studies at least part of the LTRs are missing. The actual length of the "complete" genome thus differs between studies, although all approaches cover the entire protein-coding region. Manual extraction was done in five out of eight studies, two others reported machine isolation only, while an eighth study used a plethora of isolation methods. Manual extraction obviates the need for a machine, but it can be labour-intensive and slow when dealing with a large numbers of samples. From the present study we conclude that manual RNA isolation results in a better RNA quality necessary for complete HIV-1 genome sequencing and that shorter amplicons are more successful for sequencing directly from clinical samples.

5. Conclusions

Manual isolation of HIV-1 RNA from plasma with the QIAamp Viral RNA Mini Kit yields the best results when analysing the number of amplicons generated in subsequent RT-PCR reactions for complete genome high-throughput sequencing, compared with either machine RNA isolation using the same kit or other kit/machine combinations from different suppliers. High sample age and low initial plasma viral load were found to negatively influence the amplification of HIV-1 fragments >3 kb.

Funding

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.virusres.2016.08.004.

References

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 (4), 868–882.

Bezemer, D., Cori, A., Ratmann, O., van Sighem, A., Hermanides, H.S., Dutilh, B.E., Gras, L., Rodrigues Faria, N., van den Hengel, R., Duits, A.J., Reiss, P., de Wolf, F., Fraser, C., 2015. Dispersion of the HIV-1 epidemic in men who have sex with men in the Netherlands: a combined mathematical model and phylogenetic analysis. PLoS Med. 12 (11), e1001898.

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.

Delwart, E.L., Herring, B., Rodrigo, A.G., Mullins, J.I., 1995. Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. PCR Methods Appl. 4 (5), S202–S216.

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 (12), 3838–3844.

Gall, A., Morris, C., Kellam, P., Berry, N., 2014. Complete genome sequence of the WHO International Standard for HIV-1 RNA determined by deep sequencing. Genome Announc. 2 (1).

Giallonardo, F.D., Topfer, A., Rey, M., Prabhakaran, S., Duport, Y., Leemann, C., Schmutz, S., Campbell, N.K., Joos, B., Lecca, M.R., Patrignani, A., Daumer, M., Beisel, C., Rusert, P., Trkola, A., Gunthard, H.F., Roth, V., Beerenwinkel, N., Metzner, K.J., 2014. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. Nucleic Acids Res. 42 (14), e115.

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., Gunthard, 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 (3), e1002529.

Hunt, M., Gall, A., Ong, S.H., Brener, J., Ferns, B., Goulder, P., Nastouli, E., Keane, J.A., Kellam, P., Otto, T.D., 2015. IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics 31 (14), 2374–2376.

Luk, K.C., Berg, M.G., Naccache, S.N., Kabre, B., Federman, S., Mbanya, D., Kaptue, L., Chiu, C.Y., Brennan, C.A., Hackett Jr., J., 2015. Utility of metagenomic next-generation sequencing for characterization of HIV and human pegivirus diversity. PLoS One 10 (11), e0141723.

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.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J., 2008. A large genome center's improvements to the Illumina sequencing system. Nat. Methods 5 (12), 1005–1010.

Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P., Oyola, S.O., 2012. Optimal enzymes for amplifying sequencing libraries. Nat. Methods 9 (1), 10–11.

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 (18), e144.

van der Kuyl, A.C., Bakker, M., Jurriaans, S., Back, N.K., Pasternak, A.O., Cornelissen, M., Berkhout, B., 2013. Translational HIV-1 research: from routine diagnostics to new virology insights in Amsterdam, the Netherlands during 1983–2013. Retrovirology 10, 93.

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.