Virus Research 239 (2017) 10–16

Contents lists available at ScienceDirect

Virus Research

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

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

Article history:
Received 15 June 2016
Received in revised form 2 August 2016
Accepted 4 August 2016
Available online 4 August 2016

Keywords:
HIV-1
Complete genome
RNA isolation
QIAamp viral isolation kit
High-throughput deep sequencing

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.

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

* Corresponding author.
E-mail address: a.c.vanderkuyl@amc.uva.nl (A.C. van der Kuyl).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.virusres.2016.08.004

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 °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