Journal of Virological Methods 236 (2016) 98–104

Contents lists available at ScienceDirect

Journal of Virological Methods

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

Multiplexed next-generation sequencing and *de novo* assembly to obtain near full-length HIV-1 genome from plasma virus

Shambhu G. Aralaguppe[a], Abu Bakar Siddik[a,1], Ashokkumar Manickam[b,1], Anoop T. Ambikan[c], Milner M. Kumar[c], Sunjay Jude Fernandes[d], Wondwossen Amogne[e], Dhinoth K. Bangaruswamy[c], Luke Elizabeth Hanna[b], Anders Sonnerborg[a,f], Ujjwal Neogi[a,*]

[a] *Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden*
[b] *HIV/AIDS Division, Department of Clinical Research, National Institute for Research in Tuberculosis, Indian Council of Medical Research, Chennai, India*
[c] *SciGenom Lab Pvt. Ltd, Cochin, India*
[d] *Unit of Computational Medicine, Center for Molecular Medicine, Department of Medicine & Science for Life Laboratories, Karolinska Institutet, Stockholm, Sweden*
[e] *Department of Internal Medicine, School of Medicine, Addis Ababa University, Addis Ababa, Ethiopia*
[f] *Department of Infectious Diseases, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden*

Article history:
Received 4 June 2016
Received in revised form 8 July 2016
Accepted 9 July 2016
Available online 19 July 2016

Keywords:
HIV-NFLG
De novo assembly
Plasma

A B S T R A C T

Analysing the HIV-1 near full-length genome (HIV-NFLG) facilitates new understanding into the diversity of virus population dynamics at individual or population level. In this study we developed a simple but high-throughput next generation sequencing (NGS) protocol for HIV-NFLG using clinical specimens and validated the method against an external quality control (EQC) panel. Clinical specimens (n = 105) were obtained from three cohorts from two highly conserved HIV-1C epidemics (India and Ethiopia) and one diverse epidemic (Sweden). Additionally an EQC panel (n = 10) was used to validate the protocol. HIV-NFLG was performed amplifying the HIV-genome (*Gag-to-nef*) in two fragments. NGS was performed using the Illumina HiSeq2500 after multiplexing 24 samples, followed by *de novo* assembly in Iterative Virus Assembler or VICUNA. Subtyping was carried out using several bioinformatics tools. Amplification of HIV-NFLG has 90% (95/105) success-rate in clinical specimens. NGS was successful in all clinical specimens (n = 45) and EQA samples (n = 10) attempted. The mean error for mutations for the EQC panel viruses were <1%. Subtyping identified two as A1C recombinant. Our results demonstrate the feasibility of a simple NGS-based HIV-NFLG that can potentially be used in the molecular surveillance for effective identification of subtypes and transmission clusters for operational public health intervention.

1. Introduction

One of the most prominent characteristics of HIV is the incredible genetic diversity due to its error prone reverse transcriptase, high replication rate and selection pressure exerted by the host and antiretrovirals. The constant genetic evolution within the host results in viral quasispecies (Perelson et al., 1996). High genetic diversity not only presents major challenges for vaccine development but also foremost risk for effective diagnostics and monitoring assays.

Population-based Sanger sequencing has been the standard method for clinical HIV-1 genotypic assays like genotypic resistance testing (GRT) or tropism testing (GTT). Moreover, Sanger-based whole HIV-1 genome sequencing and phylogenetic analyses have also become instrumental for efficient characterization of viruses as well as transmission clusters in an epidemic. Several studies have developed protocols for HIV-1 near full-length genome (HIV-NFLG) sequencing from cultured cells or proviral DNA from blood mononuclear cells (Lole et al., 1999; Salminen et al., 1995; Sanabani et al., 2011). Data obtained directly from plasma viruses are scarce, albeit they are routinely used in clinical HIV-1 genotypic assays (Kemal et al., 2009; Nadai et al., 2008; Sanchez et al., 2014). Recently we developed a subtype-independent HIV-NFLG amplification with two large amplicons of 5.5 kb and 3.7 kb

* Corresponding author.
E-mail address: ujjwal.neogi@ki.se (U. Neogi).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jviromet.2016.07.010
0166-0934/

followed by Sanger sequencing with 17 primers to obtain HIV-NFLG (Grossmann et al., 2015).

Next generation sequencing (NGS) technologies are increasingly being integrated into clinical practice (Curnutte et al., 2014). Further, the recent evolution and continuous improvement of the NGS technologies provides unprecedented promises for the large-scale sequencing of the virus genome in a simpler, cost and time efficient manner. Though all the NGS platforms generate enormous amount of comparable data, the key differences are in the quality of the data. Illumina platforms have simpler laboratory workflows, reduced hands-on time, and the highest throughput with the lowest systemic error rate (~0.1%) (Loman et al., 2012; Quinones-Mateu et al., 2014). The advantage of NGS has also been demonstrated for a variety of applications related to HIV-1 clinical management. It can effectively identify drug resistance mutations (DRMs) in the minor viral quasispecies (Ekici et al., 2014; Gibson et al., 2014) predicting co-receptor tropism (Archer et al., 2012) or tracking the immune escape induced by host selection pressure (Henn et al., 2012). It has also been applied in generation of HIV-NFLG (Berg et al., 2016; Gall et al., 2012). However, given the extensive heterogeneity and past failure with other HIV-genotypic tests (Aghokeng et al., 2011) further advancement and adaptation to newer techniques are important.

In this study we have developed a simple high throughput next generation sequencing protocol to generate HIV-NFLG from plasma after multiplexing several patient samples together. We also evaluated the method using an external quality control panel and molecular plasmid to examine analytical biases and intrinsic errors in sequence assembly. The method was applied on three cohorts from settings with highly conserved HIV-1C epidemics (India and Ethiopia) and from the highly diverse Swedish HIV-1 epidemic in order to describe the subtypes and investigate the genetic diversity of HIV-1.

2. Materials and methods

2.1. Virus and plasmids

To examine analytical biases and intrinsic errors in sequence reads we used nine viruses (QC panel herein) obtained from either the NIH AIDS Reagent Program, NIH, US (97ZA009, 93IN101 and 94KE105) or the EQAPOL Viral Diversity program, Duke University, US (DEMB11US015.S1, DEMB11US006.S1, DEMC11ZM006.S1, DEMC09ZA001.S1, DEMC07ZA011.S1, DEMC08NG001.S1). Additionally to examine PCR induced errors and errors from the NGS run; we used a HIV-1C infectious molecular clone, MJ4 (Ndung'u et al., 2001). 293 T cells were transfected with 3 μg of MJ4 plasmid using Fugene HD reagent (Promega, US). Seventy-two hrs. post transfection the culture supernatant (vMJ4 herein) was collected and stored. The MJ4 plasmid was also cut with *Sac*I and *Eco*RI to obtain 9.9 kb fragment (pMJ4 herein) and used directly for NGS.

2.2. Clinical specimens

Plasma samples (n = 105) were obtained from three cohorts: Swedish InfCare cohort (n = 65), Ethiopian HIV-1C cohort (n = 29) and Indian HIV-1C cohort (n = 11). The Swedish and Ethiopian samples were processed and analysed at Karolinska Institutet, Sweden while the Indian samples were processed at the National Institute for Research in Tuberculosis, Chennai, India. All the samples are collected from therapy naïve individuals with viral load >3000 copies/mL(range 3060–6600000 copies/mL)

2.3. Development of the amplification strategy

To identify the sensitivity and specificity of the method, in total 10 QC panel and 105 plasma samples were used. Viral RNA was extracted from 140 μl of plasma input using QIAamp Viral RNA Mini Kit (Qiagen, US). We applied our earlier developed protocol for HIV-NFLG with modifications with lower detection level for successful amplification of 3000 copies/mL (Grossmann et al., 2015). The HIV-NFLG was performed after amplifying the 9 kb HIV-genome in two fragments (F1-*Gag-to-vpu* and F2-*Tat-to-3LTR*). All PCR was reduced to 20 cycles for KAPA HiFi Taq DNA polymerase as per manufacturer's protocol.

2.4. Frequency of in vitro recombination

To find out the frequency of recombination due to PCR, we have mixed DEMB11US015.S1 and DEMC09ZA001.S1 samples with 1:1, 1:4 and 4:1 based on the viral copy number (9.01×10^9 copies/mL and 3.94×10^9 copies/mL respectively) cDNA for F1-*Gag-to-vpu* was converted using gene specific primer. cDNA were further endpoint diluted (dilution factor of 10 and upto 6th dilutions; 10^{-6}) and F1-*Gag-to-vpu* were amplified as described previously (Grossmann et al., 2015). As Gag gene was reported to be one of the major hot spot for recombination (Smyth et al., 2014), we sequenced the complete Gag region.

2.5. NGS using illumina HiSeq2500

For NGS, we selected 10 culture supernatants and 45 plasma samples. Both amplified fragments (F1-*Gag-to-vpu* and F2-*Tat-to-3LTR*) were gel-purified using PureLink Quick Gel Extraction Kit (Invitrogen, USA) and mixed in equimolar concentration (10 nM each). The mix was then fragmented on the Covaris S200 at 300bp for 75 s with peak power- 50 and cycle/burst -200. The library was prepared using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolab, US) with multiplexed NEB next adaptors. The samples were then pooled together (either 24 or 10) along with other unrelated non-viral indexed libraries. Paired end sequencing of length 250 bp was carried out on the Illumina HiSeq2500. Of the 24 samples run there was at least one culture supernatant as QC sample. To define the reproducibility, 10 samples were also re-sequenced. We totally perform 2×24 samples and 2×10 samples run.

2.6. HIV-NFLG de novo assembly

The generated sequences were de-multiplexed and adapter trimmed using Cutadapt v1.8 program using the default setting (error rate <0.1) (Martin, 2011) followed by removal of the low quality bases (phred value score <Q30) by Sickle ver 1.33 (Joshi and Fass, 2011). Duplicate reads were removed using FastUniq (Xu et al., 2012). The *de novo* genome assembly was performed by IVA (Iterative Virus Assembler) (Hunt et al., 2015). For any sample, when IVA was not able to yield a single contig of length ~9KB, we used VICUNA (Yang et al., 2012). Then the filtered sequences were re-aligned against individual NFLG in order to estimate 'position specific base count' which, was performed using the 'count' program available in IGV (Robinson et al., 2011). The consensus sequences were generated from the base-count data with occurrence of a particular nucleotide >50% in a given position and used for subtyping and phylogenetic analysis. Circos was used to visualize the genome diversity (Krzywinski et al., 2009). The scheme and bioinformatics pipeline is presented in Fig. 1.

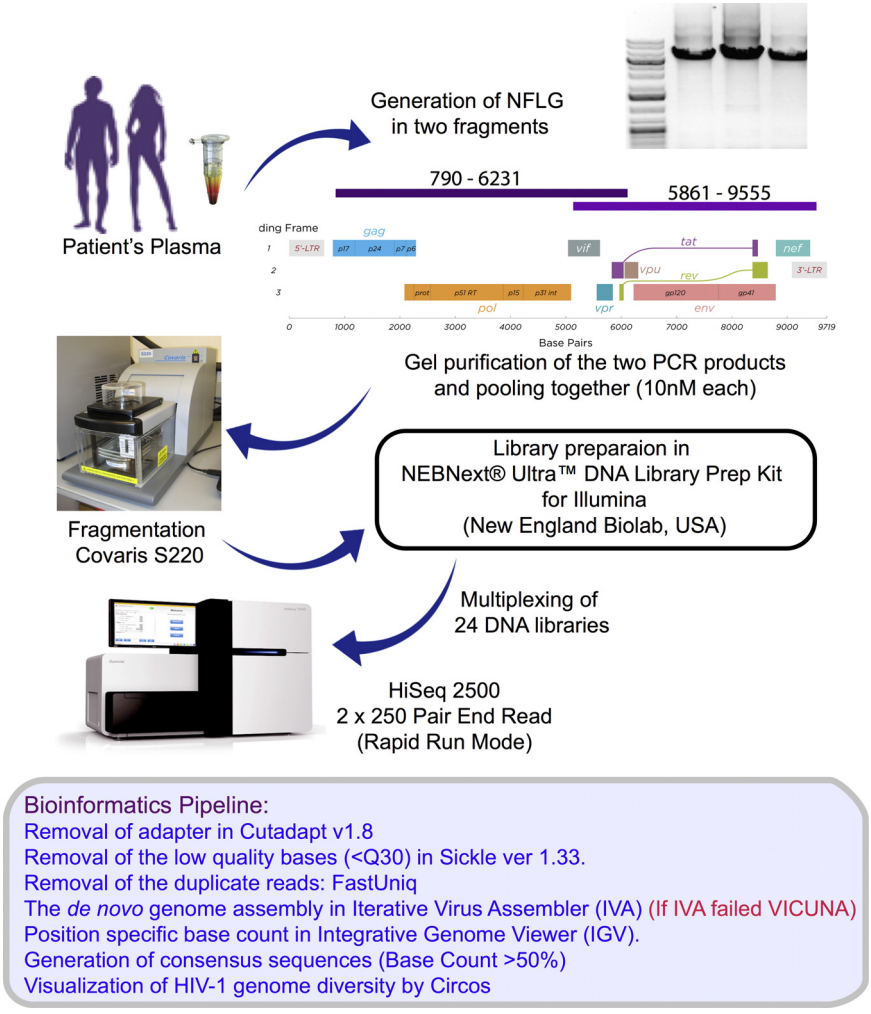

Fig. 1. Scheme of HIV-NFLG using NGS and bioinformatics pipeline.

2.7. *HIV-NFLG by Sanger sequencing*

To validate the consensus assembly generated from NGS (Henn et al., 2012), we used a subset of samples (n = 20) for generation of HIV-NFLG by Sanger sequencing as described previously (Grossmann et al., 2015). The contig was generated using the ContigExpress in Vector NTI Express v1.1.5 (Life Technologies, US).

2.8. *Subtyping, phylogenetic analysis and recombination analysis*

Subtyping was performed using Rega v3 (Pineda-Pena et al., 2013) and COMET HIV-1 (Struck et al., 2014) followed by HIV-1 jpHMM (Schultz et al., 2009). Maximum likelihood phylogenetic analysis was performed in Fast tree v3 with general time-reversible nucleotide substitution model with inverse gamma distribution (Price et al., 2010). The recombination was further identified in RDP2 (Martin et al., 2005) and bootscan analysis in SimPlot Version 3.5.1 (Lole et al., 1999).

2.9. *Ethical considerations*

Ethical permissions were obtained from the respective sites. Swedish samples: Regional Ethics Committee Stockholm (Dnr: 2006/1367-31/4). Ethiopian samples: the Ethiopian Science and Technology Agency (Ref. No. RPHE/126-83/08), and the Drug Administration and Control, Authority of Ethiopia (Ref. No. 02/6/22/17). Indian samples: National Institute for Research in Tuberculosis Institutional Ethics Committee (NIRT IEC No: 2009009). All participants gave written informed consent. The patient information was anonymized and de-linked prior to analysis.

3. Results

3.1. *Amplification efficiency, sensitivity and specificity*

As a first step we used 10 QC panel viruses and 105 plasma samples with viral load >3000 copies/mL from the three cohorts. From the QC panel, we successfully amplified both fragments from all viruses. Among the 105 plasma samples we were able of amplify both fragments from 95 samples (90% success rate). In the remaining 10 samples, F1 did not work for eight samples while F2 did not work for one sample. In one sample both fragments didn't work. For specificity, we used 20 HIV negative clinical specimen, which didn't show any amplification, giving a specificity of 100% towards HIV-1 positive samples.

3.2. *PCR induced* in vitro *recombination*

We have mixed HIV-1B (DEMB11US015.S1) and HIV-1C (DEMC09ZA001.S1) viruses in 1:1, 1:4 and 4:1 dilutions with dilution factor 10 and performed limited dilutions upto 10^{-6} dilutions

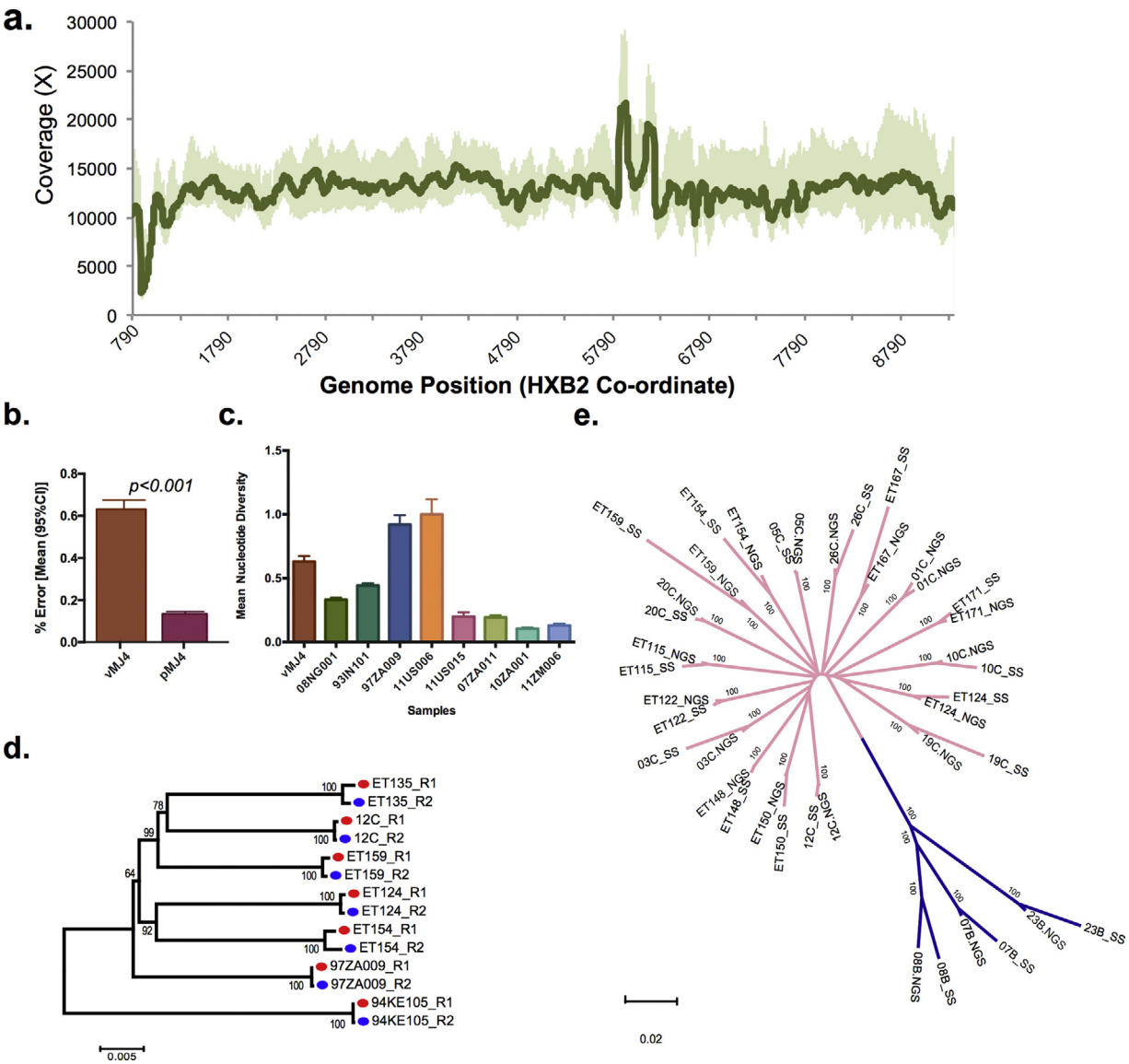

Fig. 2. NGS coverage, intrinsic PCR induced error and reproducibility: (A) The median (IQR) coverage was 13023X (IQR: 10390-16910) with 100% coverage of the HXB2 position 790–9417 (*gag-to-nef*). (B) The mean error calculated using pMJ4 and vMJ4 were 0.13% vs 0.63%; ($p < 0.001$) which is less than 1%. (C) The mean nucleotide diversity per position at the complete coverage depth based on the sequence available at the GenBank of the QC panel viruses. The obtained mean diversity is <1% per position. (D) The maximum likely (ML) phylogenetic analysis of the consensus sequences obtained in two separate runs, identified clustering of the same samples together with 100 bootstrap supports. (E) Concordance between consensus sequences generated from the NGS and the Sanger sequencing. The ML phylogenetic analysis identified clustering of the consensus sequences generated by NGS and Sanger sequencing together with 100% bootstrap support from a sample.

and sequenced *gag* gene from all amplified fragments. In 1:4 (B:C) all the sequences were identified as HIV-1C while in 4:1 (B:C) all the sequences were identified as HIV-1B. In 1:1 dilution samples No inter-subtype recombinant was identified in the gag region, which is more prone to recombination (Supplementary Fig. 1)

3.3. *NGS and* de novo *assembly success rate*

We attempted 10 QC panel viruses and 45 plasma samples and successfully obtained the NFLG in all panel viruses and samples. *De novo* assembly was successful for 49 samples using IVA while for the remaining six samples VICUNA was used as IVA failed to generate the NFLG contig. We also compared the contig generated by IVA and VICUNA of the QC panel viruses and observed ~97% nucleotide identity by both methods. The median (IQR) coverage depth was 13,023X (IQR: 10,390–16,910) with 100% coverage of HXB2 position 790–9417 (*gag-to-nef*) (Fig. 2A). Except sample DEMC07ZA011.S1 (95% identity), all the remaining samples gave >99% identity with the sequences available in the GenBank.

3.4. *PCR induced errors and errors generated by NGS*

To check PCR induced errors and errors due to NGS we used two strategies. First the plasmid pMJ4 was digested with *Sac*I and *Eco*RI to obtain 9 kb full length HIV-1 genome. Second, we amplified two fragments from the RNA extracted from the virus following transfection of the plasmid. The mean error calculated based on the pMJ4 and vMJ4 were <1% (0.13% vs 0.63%; $p < 0.001$) (Fig. 2B). Total of eight nucleotide miss-matches were observed between the MJ4 sequences (GenBank Acc No. AF321523) compared to pMJ4 and vMJ4. Between pMJ4 and vMJ4 only one nucleotide mismatch was observed in the consensus sequences. We also calculated the mean nucleotide diversity per position at the complete coverage depth based on the sequence available at the GenBank of the QC panel

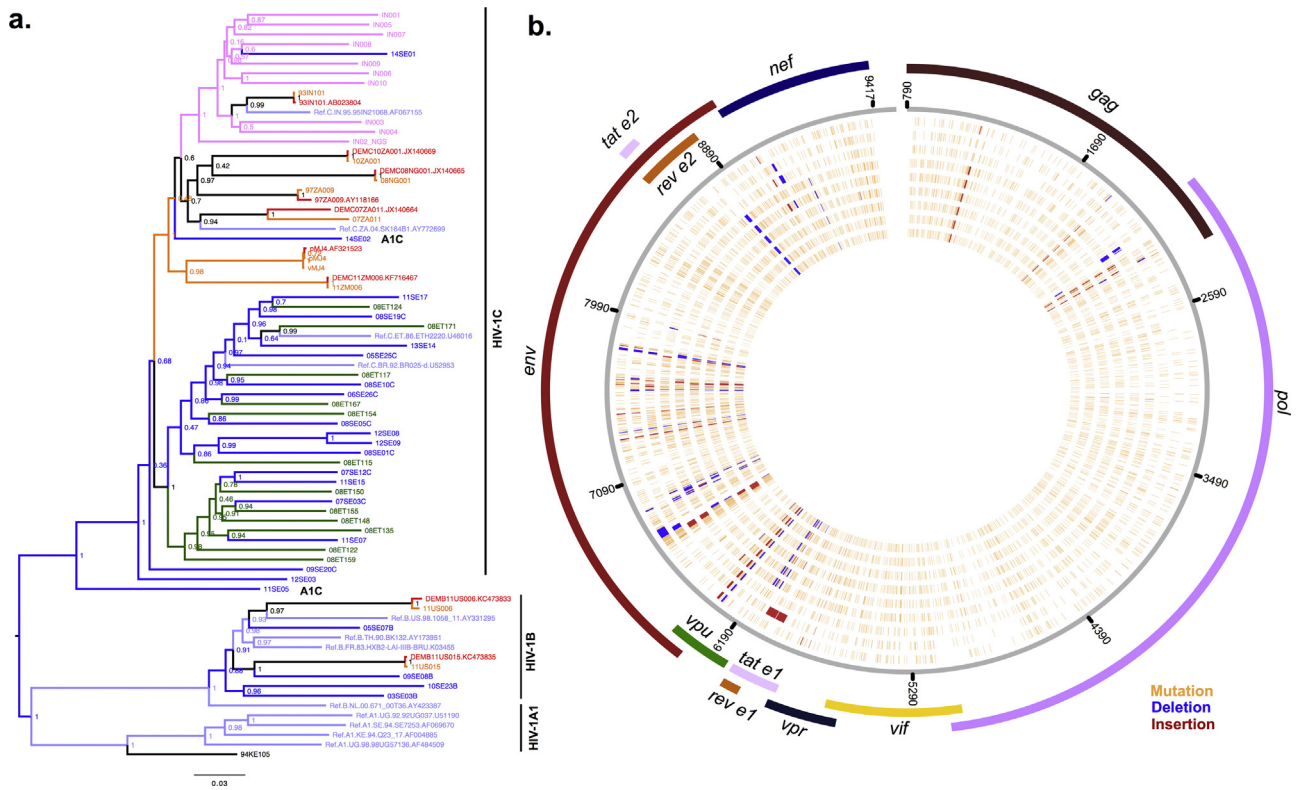

Fig. 3. Phylogenetic analysis and genetic diversity of HIV-1 subtypes: (A) Maximum likelihood phylogenetic analysis of the consensus HIV-NFLG obtained by NGS from Indian HIV-1C cohort (pink), Ethiopian HIV-1C cohort (green) and Swedish cohort (blue). Sequenced EQC panel sequences (orange) clustered with the original sequences (red) obtained from the database. (B) Genetic diversity of HIV-1 genes at NFLG-level. Nine representative HIV-1 genome sequences from HIV-1B (outer three) and HIV-1C (inner six) from three cohorts are shown in circular format. The sequences are compared to the HXB2 sequences (Acc No K03455) (shown in grey). Nine HIV-1 genes are presented as per open reading frame based on HXB2 co-ordinates. Single nucleotide polymorphisms are shown in yellow, insertion in red and deletions in blue.

viruses and obtained mean diversity <1% per position (Fig. 2C). This indicates very low level of PCR induced errors and errors due to NGS. The mean unmapped read on the contig generated by IVA or VICUNA after cleaning the data is as low as 2.61% (Supplementary Fig. 2).

3.5. NGS reproducibility and HiSeq2500 and Sanger sequence concordance

To identify the reproducibility 10 samples were run in two separate runs, which identified >99% similarity in all the samples. The phylogenetic analysis revealed clustering of the same samples together with 100 bootstrap supports (Fig. 2D). We observed a high concordance in nucleotide identity between the consensus sequences generated from NGS and Sanger sequencing. The phylogenetic analysis identified clustering of the consensus sequences generated by NGS and Sanger sequencing with 100% bootstrap support from a sample (Fig. 2E).

3.6. HIV-1 subtyping and genetic diversity

HIV-1 subtyping by automated tools and maximum likelihood phylogenetic analysis identified 87% (39/45) as HIV-1C, 8% (4/45) as HIV-1B and the remaining two samples were identified as A1C recombinants (Fig. 3A). The mosaic pattern of the A1C recombinant is presented as a Supplementary Fig. 3. It is to be noted that *pol*-gene subtyping that is mostly used for subtyping described these recombinant strains as HIV-1C. We also compared and visualized the genetic diversity plot, observing cluster and subtype specific variations across the genome. As expected, inter-subtype variability was observed more in *env* gene while *pol* gene was more conserved. Representative genome sequences from HIV-1B and HIV-1C are presented in Fig. 3B.

4. Discussion

In this study we demonstrate the feasibility of a high-throughput next generation sequencing protocol to assemble near full-length HIV-1 genome from plasma virus that is used as a source of most routine HIV-1 genotypic tests. The protocol is simple and labour-efficient with extremely high-throughput (>10,000X coverage per nucleotide). It showed improved detection of recombinant forms and can also be more effective in identifying phylogenetic transmission clusters in a local HIV-1 epidemic (Novitsky et al., 2015).

Sequence-based genotyping assays in HIV are always challenging and subject to subtype-specific constraints, especially when dealing with HIV-1 non-B viruses, which predominate in low- and middle-income countries (LMICs) (Aghokeng et al., 2011). Even the US Food and Drug Administration (FDA) approved commercial assays like ViroSeq HIV-1 Genotyping System (Abbott, US) failed to perform efficiently in non-B subtype dominated epidemics like in Cameroon (Aghokeng et al., 2011) and Senegal (Thiam et al., 2013). Several countries have therefore adapted *in house* assays based on circulating subtypes in the local epidemics. Recently NFLG using NGS has been developed that amplified ~9 kb HIV-1 genome using four or six fragments followed by deep-sequencing (Gall et al., 2012; Ode et al., 2015; Zanini et al., 2015). However, the risk of amplification failure increases with increasing numbers of fragment amplifications as in the case of the ViroSeq method (sequencing primers) in diverse subtypes (Aghokeng et al., 2011; Thiam et al., 2013). Also the applicability of these methods

in diverse HIV-1 epidemics is still not evaluated. While applying the universal amplification protocol developed by Gall et al., for South African HIV-1C, three of the four fragments primers were adapted for HIV-1C which yielded 70% success rate (Danaviah et al., 2015). In our study we first developed primer sets for HIV-1C and tested them in three HIV-1C cohorts from Sweden (including mainly East Africans), Ethiopia and India, which showed >90% sensitivity. In addition, the amplification strategy is highly sensitive for HIV-1B, CRFs 01_AE and 02_AG (Grossmann et al., 2015). These subtypes together with HIV-1C cover nearly two-thirds of HIV-infected individuals globally. Recently, Berg et al. developed a new method of sequencing HIV-1 cDNA (HIV-SMART) to increase the success of HIV-NFLG; however, the complexity of the RNA extraction (in a closed m2000sp instrument) increases the operational costs and limited the applicability of the method in LMICs (Berg et al., 2016). In our current strategy the effective cost for one sample is $130–$150 if 24 samples are pooled together. The cost can be reduced further after multiplexing a higher number of samples with lower coverage requirements. The coverage can potentially be adjusted using the Lander–Waterman equation, which Illumina provides as an Excel spreadsheet for their different platforms (http://support.illumina.com/downloads/sequencing_coverage_calculator.html) (Lander and Waterman, 1988).This method for the NGS can easily be established in a central core facility centrally in a country or region. Sample processing to generate the library can be done in a routine molecular biology laboratory. The bioinformatics pipeline is also simple and can be adapted in routine NGS bioinformatics facilities.

Some technical limitations and advantages merit comments. First, the protocol is efficient only in plasma samples with a viral load >3000 copies/mL with the 140 μl input. Increasing the initial input to 1 ml of plasma followed by high-speed centrifugation may increase the sensitivity. Second, the *de novo* assembler IVA failed to generate the contig for six samples as the analysis terminated abruptly. Use of a second tool VICUNA was used on the failed samples. If both the tools failed to generate contig, other *de novo* assemblers can be used after appropriate validation. The reference-based alignment is not recommended. Finally, despite a high viral load we were not able to amplify 10 samples (mainly the F1 fragment). This could be due to mutations in the primer binding sites. However, several advantages exist. The major advantage is the broad applicability. Unlike other studies (Berg et al., 2016; Gall et al., 2012; Henn et al., 2012) we applied the method in three distinct cohorts and obtained high success in all three and it was found that amplification strategy is highly sensitive for HIV-1C, HIV-1B, 01_AE and 02_AG. The method is also less labour-intensive and simple in both wet (PCR/NGS) and dry (bioinformatics) laboratory setups. The input plasma sample volume is only 140 μl, thus it can be combined with the existing routine genotypic resistance testing without any additional blood specimen. Finally, the method is validated against the EQC panel, thus quality assured for large-scale molecular epidemiological surveillance studies.

In summary, here we demonstrate the practicality of a high-throughput NGS protocol to assemble near full-length HIV-1 genomes from plasma virus that is frequently used as the source for most routine HIV-1 genotypic tests. The method was validated in an external quality assurance panel as well as in three cohorts, obtained from both high and low income countries. The application of the method in population dynamics studies with high and efficient identification of recombinant strains and transmission clusters could enable public health based effective interventions. The method and the bioinformatics pipeline are simple and can be adapted for large-scale molecular epidemiological surveillance studies, even in LMICs.

Conflict of interest

Dhinoth Kumar Bangaruswamy, Anoop T and Milner Kumar are employee of SciGenom Pvt. Ltd. India. Other authors none to declare.

Acknowledgements

Authors would like to thank Dr. Beena PS, India, for her technical support in the NGS. The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 97ZA009, 93IN101 and 94KE105 from The UNAIDS Network for HIV Isolation and Characterization and HIV-1 MJ4 Infectious Molecular Clone (pMJ4) from Drs. Thumbi Ndung'u, Boris Renjifo and Max Essex. Authors also acknowledge the samples received through EQAPOL Viral Diversity program, Duke University, US, which has been supported in whole or part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Contract Number NOI-AI-85341. The study is partially funded by a Young Investigator Grant from Karolinska Institutet Research Foundation Grant to UN, Swedish Research Council Grant and Stockholm County Council to AS. SGA is partially supported by a doctoral funding grant (KID) from Karolinska Institutet. ABS acknowledges the scholarship received from the Swedish Institute (SI). Grant numbers and sources of support: The Karolinska Institutet Research Foundation Grant (2014fobi41250), Swedish Research Council (2012-3476), Stockholm County Council (20130042)

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.jviromet.2016.07.010.

References

Aghokeng, A.F., Mpoudi-Ngole, E., Chia, J.E., Edoul, E.M., Delaporte, E., Peeters, M., 2011. High failure rate of the ViroSeq HIV-1 genotyping system for drug resistance testing in Cameroon, a country with broad HIV-1 genetic diversity. J. Clin. Microbiol. 49, 1635–1641.

Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., Paxinos, E., Arts, E.J., Robertson, D.L., Mimms, L., Quinones-Mateu, M.E., 2012. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One 7, e49602.

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.

Curnutte, M.A., Frumovitz, K.L., Bollinger, J.M., McGuire, A.L., Kaufman, D.J., 2014. Development of the clinical next-generation sequencing industry in a shifting policy climate. Nat. Biotechnol. 32, 980–982.

Danaviah, S., Manasa, J., Wilkinson, E., Pillay, S., Sibisi, Z., Msweli, S., Pillay, D., deOliveira, T., 2015. Near full-length HIV-1 sequencing to understand HIV phylodynamics in Africa in real time. Conference on Retroviruses and Opportunistic Infections (CROI).

Ekici, H., Rao, S.D., Sonnerborg, A., Ramprasad, V.L., Gupta, R., Neogi, U., 2014. Cost-efficient HIV-1 drug resistance surveillance using multiplexed high-throughput amplicon sequencing: implications for use in low- and middle-income countries. J. Antimicrob. Chemother. 69, 3349–3355.

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.

Gibson, R.M., Meyer, A.M., Winner, D., Archer, J., Feyertag, F., Ruiz-Mateos, E., Leal, M., Robertson, D.L., Schmotzer, C.L., Quinones-Mateu, M.E., 2014. Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. Antimicrob. Agents Chemother. 58, 2167–2185.

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.

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, 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, 2374–2376.

Joshi, N., Fass, J., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].

Kemal, K.S., Reinis, M., Weiser, B., Burger, H., 2009. Methods for viral RNA isolation and PCR amplification for sequencing of near full-length HIV-1 genomes. Methods Mol. Biol. 485, 3–14.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645.

Lander, E.S., Waterman, M.S., 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2, 231–239.

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.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. 30, 434–439.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 10–12.

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.

Ndung'u, T., Renjifo, B., Essex, M., 2001. Construction and analysis of an infectious human Immunodeficiency virus type 1 subtype C molecular clone. J. Virol. 75, 4964–4972.

Novitsky, V., Moyo, S., Lei, Q., DeGruttola, V., Essex, M., 2015. Importance of viral sequence length and number of variable and informative sites in analysis of HIV clustering. AIDS Res. Hum. Retroviruses.

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.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., Ho, D.D., 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271, 1582–1586.

Pineda-Pena, A.C., Faria, N.R., Imbrechts, S., Libin, P., Abecasis, A.B., Deforche, K., Gomez-Lopez, A., Camacho, R.J., de Oliveira, T., Vandamme, A.M., 2013. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infect. Genet. Evol. 19, 337–348.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Quinones-Mateu, M.E., Avila, S., Reyes-Teran, G., Martinez, M.A., 2014. Deep sequencing: becoming a critical tool in clinical virology. J. Clin. Virol. 61, 9–19.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Salminen, M.O., Koch, C., Sanders-Buell, E., Ehrenberg, P.K., Michael, N.L., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. Virology 213, 80–86.

Sanabani, S.S., Pastena, E.R., da Costa, A.C., Martinez, V.P., Kleine-Neto, W., de Oliveira, A.C., Sauer, M.M., Bassichetto, K.C., Oliveira, S.M., Tomiyama, H.T., Sabino, E.C., Kallas, E.G., 2011. Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in Sao Paulo, Brazil. PLoS One 6, e25869.

Sanchez, A.M., DeMarco, C.T., Hora, B., Keinonen, S., Chen, Y., Brinkley, C., Stone, M., Tobler, L., Keating, S., Schito, M., Busch, M.P., Gao, F., Denny, T.N., 2014. Development of a contemporary globally diverse HIV viral panel by the EQAPOL program. J. Immunol. Methods 409, 117–130.

Schultz, A.K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., Stanke, M., 2009. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res. 37, W647–51.

Smyth, R.P., Schlub, T.E., Grimm, A.J., Waugh, C., Ellenberg, P., Chopra, A., Mallal, S., Cromer, D., Mak, J., Davenport, M.P., 2014. Identifying recombination hot spots in the HIV-1 genome. J. Virol. 88, 2891–2902.

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.

Thiam, M., Diop-Ndiaye, H., Kebe, K., Vidal, N., Diakhate-Lo, R., Diouara, A.A., Leye, N., Ndiaye, O., Sow, A., Ngom-Gueye, N.F., Mboup, S., Toure-Kane, C., 2013. Performance of the ViroSeq HIV-1 genotyping system v2.0 on HIV-1 strains circulating in Senegal. J. Virol. Methods 188, 97–103.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., Chen, S., 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7, e52249.

Yang, X., Charlebois, P., Gnerre, S., Coole, M.G., Lennon, N.J., Levin, J.Z., Qu, J., Ryan, E.M., Zody, M.C., Henn, M.R., 2012. De novo assembly of highly diverse viral populations. BMC Genomics 13, 475.

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.