Journal of Virological Methods 187 (2013) 314–320

Contents lists available at SciVerse ScienceDirect

Journal of Virological Methods

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

Low abundance drug resistance variants in transmitted HIV drug resistance surveillance specimens identified using tagged pooled pyrosequencing

Hezhao Ji[a], Ben Liang[b], Yang Li[a], Gary Van Domselaar[b], Morag Graham[c,d], Shaun Tyler[c], Harriet Merks[a], Paul Sandstrom[a], James Brooks[a,*]

[a] National HIV & Retrovirology Laboratories, National Microbiology Laboratory, Public Health Agency of Canada, Ottawa, Canada
[b] Bioinformatics Core, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada
[c] Genomics Core, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada
[d] Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada

Article history:
Received 16 May 2012
Received in revised form 10 October 2012
Accepted 22 October 2012
Available online 14 November 2012

Keywords:
HIV-1
Transmitted drug resistance
Surveillance
Tagged pooled pyrosequencing
Sanger sequencing

A B S T R A C T

HIV drug resistance (DR) testing using Sanger sequencing (SS) is limited by the inability of the method to identify low abundance drug resistance variants. The application of tagged pooled pyrosequencing (TPP) for HIV DR surveillance is described and the results compared with SS. HIV$^+$ serum specimens were genotyped using both SS and TPP. Surveillance drug resistance mutations were identified using SS and TPP consensus reads at multiple mixed base identification thresholds (MBITs). Drug resistance patterns were highly concordant between SS and TPP when the MBIT was set at 20%. DR mutations were detected in 7.1% of the subjects, with 1.6% of individuals harboring resistance to NRTI, 3.3% NNRTI and 2.7% PI. Analyzing the TPP reads for each subject confirmed that drug resistance mutations with frequencies <20% were inconsistently detected by SS. Conversely, low abundance drug resistant variants were easily identified using TPP with mixed base identification threshold set at low value. In conclusion, at considerable savings when compared to commercial assays, TPP produces HIV DR profiles that are concordant with those from SS, furthermore, these same data can be used to identify low abundance drug resistant variants.

1. Introduction

Antiretroviral therapy has significantly reduced the morbidity and mortality due to HIV/AIDS (Burger et al., 2004; Palella, Jr. et al., 1998; Walensky et al., 2006), however treatment can fail after the emergence of drug resistant viral strains (Bartlett et al., 2001; Booth and Geretti, 2007; Public Health Agency of Canada, 2009). Surveillance for transmitted HIV drug resistance (HIV TDR) remains an important component in effective HIV/AIDS management strategies (Bennett et al., 2008a, 2008b) especially in the context of treatment as prevention efforts (Van and Boucher, 2010). These data facilitate optimal empiric treatment choices; inform population-level recommendations on HIV treatment (Bennett, 2006; Bennett et al., 2008a, 2009); and provide an index of the effectiveness of ongoing HIV management strategies (De et al., 2004; Oette et al., 2006; Van et al., 2002).

Conventional HIV DR testing employs Sanger sequencing (SS) of the HIV *pol* gene which is then interrogated for the presence of DR mutations. Although SS-based genotyping is the most popular method for HIV TDR surveillance test, there are inherent limitations to the methodology. For example, interpretation of mixed base calls is often subjective and the threshold at which these mixtures are identified is usually higher (Demeter et al., 1998; Schuurman et al., 1999) than that which has shown to be clinically relevant (Halvas et al., 2010; Johnson et al., 2008; Metzner et al., 2009; Paredes et al., 2010; Simen et al., 2009). Hence, it would be beneficial to establish a cost-effective approach that detects low abundance drug resistance variants (LADRV) routinely for HIV TDR surveillance or clinical management.

The commercial launch of 454 pyrosequencing system in 2005 marked the advent of the era of next generation sequencing. 454 pyrosequencing provides massive parallel sequencing producing an array of clonal reads from mixed sequences (Margulies et al., 2005; Shendure and Ji, 2008). The current applications of pyrosequencing have been to, either resolve the sequences from different organisms within a biome (Andersson et al., 2008; Zhang et al., 2009) or probe low frequency variants existing within a single target (Hoffmann et al., 2007; Rozera et al., 2009; Simen et al., 2009; Wang et al., 2007). While there has been a great deal of interest in the application of pyrosequencing to address questions oriented around HIV genetic diversity within an individual; it has previously demonstrated that pooled pyrosequencing has great potential to

* Corresponding author at: National HIV & Retrovirology Laboratories, Public Health Agency of Canada, Room 3172, 100 Eglantine Driveway, Ottawa, ON, Canada K1A 0K9. Tel.: +1 613 946 0120; fax: +1 613 957 1163.
E-mail address: james_brooks@phac-aspc.gc.ca (J. Brooks).

0166-0934/$ – see front matter.
http://dx.doi.org/10.1016/j.jviromet.2012.10.018

be applied to examine virus diversity within a population (Ji et al., 2010). This report also highlighted some of the challenges of unlabeled pooling such as incomplete coverage of the *pol* gene and the inability of associating reads to specific subjects. In a subsequent study, it was demonstrated that the use of barcoded primers on dried blood spot specimens allowed us to address these issues and resolve TDR patterns, from both the population and from an individual, in a method so-called tagged pooled pyrosequencing (TPP) (Ji et al., 2011). The current report describes the application of TPP to genotype serum specimens for routine TDR surveillance with the specific advantage of being able to resolve LADRV within individual specimens at no additional cost.

2. Materials and methods

2.1. Subjects and specimens

Serum specimens were collected from 183 anonymized, treatment-naïve HIV-1 infected subjects in Canada for Research Ethics Board exempt routine public health surveillance. Total nucleic acid was extracted from 200 μl of serum using the Nuclisens EasyMag system (Biomerieux, Saint-Laurent, Canada) following manufacturer's instructions. All extracted specimens were then genotyped by both a well-optimized, SS-based in-house assay and TPP.

2.2. HIV-1 genotyping by SS and TPP

For SS-based HIV genotyping, HIV-1 protease and reverse transcriptase, up to codon 236, were sequenced bi-directionally with an in-house protocol as described previously (Ji et al., 2010). In brief, each nucleic acid extract was reverse transcribed, amplified with nested PCR and sequenced using ABI Prism BigDye 3.1 on an ABI 3130XL (Applied Biosystems, USA). The detailed SS workflow, the sequences of primers applied and the PCR cycling conditions are included in supplement document (Supplement 1).

For TPP, nucleic acid extraction, reverse transcription and one round of PCR were performed as above. The nested-PCR was conducted using fusion primers with all forward primers tagged by designated multiplex identifiers (MID), generating three overlapping amplicons pre-labeled using the same MID for each subject (454 Sequencing Technical Bulletin No.005-2009, 2009; 454 Sequencing Technical Bulletin No.013-2009, 2009). All MID tagged forward fusion primers consisted four parts: 454 primer A adaptor sequence (5′-CGTATCGCCTCCCTCGCGCCA-3′), 4-base key (TCAG), a suitable 10-base MID sequence, and the HIV gene specific sequence located at the 3' terminal. In contrast, all reverse fusion primers consisted only three parts: 454 primer B adaptor (5′-CTATGCGCCTTGCCAGCCCGC-3′), the same key sequences and the HIV specific sequence. The gene-specific sequences of the applied fusion primers were: TPP-1F: 5′-TCCCTCARATCACTCTTTGG-3′, TPP-1R: 5′-GGRTTTTYAGGCCCAATTTT-3′, TPP-2F: 5′-TKAAAGCCAGGRATGGATGG-3′; TPP-2R: 5′-TCCCTGGTGTCTCATTGTTT-3′; TPP-3F: 5′-AGTACTRGATGTGGGWGATGC-3′; TPP-3R: 5′-CTGTCCATTTRTCAGGATGRA-3′. All nested PCR procedures were performed using common reaction conditions at annealing temperature of 58.9 °C (Supplement 1). The three overlapping PCR amplicons for each specimen were purified, quantified and pooled at equimolar concentrations using a Beckman Biomex FX system equipped with a DTX880 Multimode Detector (Beckman, Boulevard Brea, USA). Individual pooled amplicon mixtures were then pooled once again with mixtures from other specimens to be pyrosequenced in the same picotiterplate region. The ultimate template mixture was then clonally amplified on beads through emulsion PCR, loaded onto picotiterplate (24 specimens/lane) fitted with a 16-lane gasket and pyrosequenced using the GS FLX Titanium kit (Roche, Indianapolis, USA). An outline of the TPP workflow is shown in Fig. 1 and more TPP experimental details are provided in Supplement 1.

2.3. Sequence analysis

SS sequences were assembled and edited in Seqscape v2.5 (Applied Biosystems, Foster City, USA) with sequence variations identified by aligning the sequences with HXB-2 (Accession number: K03455). All sequence variations were confirmed through agreement of bidirectional sequencing results. Surveillance DR mutations (SDRM) were identified using the calibrated population resistance tool 4.1 (Stanford HIV database, URL: http://cpr.stanford.edu/cpr) by referring to SDRM list 2009 (Bennett et al., 2009). The overall SDRM prevalence and frequencies of all identified SDRM were then recorded for inter-group comparisons. Subtypes were determined using the REGA HIV-1 Subtyping Tool (http://www.bioafrica.net/virus-genotype/html/subtyping.html).

All TPP reads that satisfied the GS FLX default quality screening criteria were first binned according to subject-specific MIDs using the Roche Amplicon Variant Analyzer software. Individual sequence contigs were then assembled for each individual subject using HXB-2 as the reference. To further improve the reliability of the acquired TPP reads, all remaining reads were subject to a second screening with an in-house Perl script, in which reads were considered valid only if they did not contain human sequences and mapped to HXB2 with 60% overlap and ≥75% identity. The overall TPP error rate was estimated by PCR amplifying and pyrosequencing pedigreed plasmids containing HIV *pol* sequences on the same picotiterplate. To approximate mixture identification thresholds reported for SS, consensus sequences were generated for each individual using mixed base identification thresholds of 20% (MBIT20). LADRV were identified through analysis of the population of reads from an individual and represented by at different mixture thresholds of 15%, and 10% (MBIT15, MBIT10). Inter-group comparisons were conducted on SS and TPP sequences with concordance and the identification of LADRV evaluated for each specimen.

3. Results

All 183 samples were genotyped successfully using both SS and TPP methods. A total of 186,312 valid TPP reads were obtained and included in the data analysis representing an average of 1018 copies of input templates per subject were pyrosequenced successfully. The resulting average oversampling was 309 for each nucleotide position of the *pol* gene. The HIV-1 subtype distribution in the cohort was: 165 subtype B, 10 subtype C, 3 CRF01-AE, 2 CRF02-AG and one each of subtypes A1, F and G. All primers performed well with no biased amplification observed on all examined HIV subtypes suggestive of their broad serotype coverage and applicability in varied regions where different HIV serotype may be dominant.

3.1. Accuracy and reliability of TPP sequences

As sequencing errors may be introduced by PCR prior to pyrosequencing, or by the pyrosequencing process itself, a rigorous quality control process was included. To determine the overall error rate of this protocol, two pedigreed plasmids carrying HIV-1 *pol* gene were processed in parallel with all the other specimens. The results showed that the overall sequence error rate of TPP was 0.31%. Homopolymeric regions, with runs of ≥3 identical bases (Wang et al., 2007), had higher error rates (0.59%) compared with those that were classified otherwise (0.22%). Measured error rates at predicted drug resistance codons were 0.44%.