THE AGA KHAN UNIVERSITY

eCommons@AKU

Internal Medicine, East Africa

Medical College, East Africa

2016

Short Communication: Analysis of Minor Populations of Human Immunodeficiency Virus by Primer Identification and Insertion-Deletion and Carry Forward Correction Pipelines

Paul Hughes
Seattle Children's Research Institute

Wenjie Deng
University of Washington

Scott C. Olson
Seattle Children's Research Institute

Robert W. Coombs
University of Washington

Michael Chung
Aga Khan University, michael.chung@aku.edu

See next page for additional authors

Follow this and additional works at: https://ecommons.aku.edu/eastafrica_fhs_mc_intern_med

Part of the Internal Medicine Commons

Recommended Citation

Hughes, P., Deng, W., Olson, S. C., Coombs, R. W., Chung, M., Frenkel, L. M. (2016). Short Communication: Analysis of Minor Populations of Human Immunodeficiency Virus by Primer Identification and Insertion-Deletion and Carry Forward Correction Pipelines. *AIDS Research and Human Retroviruses, 32*(3), 296-302.
Available at: https://ecommons.aku.edu/eastafrica_fhs_mc_intern_med/56

Authors
Paul Hughes, Wenjie Deng, Scott C. Olson, Robert W. Coombs, Michael Chung, and Lisa M. Frenkel

This article is available at eCommons@AKU: https://ecommons.aku.edu/eastafrica_fhs_mc_intern_med/56

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 32, Number 3, 2016
© Mary Ann Liebert, Inc.
DOI: 10.1089/aid.2015.0202

Short Communication: Analysis of Minor Populations of Human Immunodeficiency Virus by Primer Identification and Insertion-Deletion and Carry Forward Correction Pipelines

Paul Hughes,[1] Wenjie Deng,[2] Scott C. Olson,[1,3] Robert W. Coombs,[4,5] Michael H. Chung,[4,5] and Lisa M. Frenkel[1,3,4,6]

Abstract

Accurate analysis of minor populations of drug-resistant HIV requires analysis of a sufficient number of viral templates. We assessed the effect of experimental conditions on the analysis of HIV *pol* 454 pyrosequences generated from plasma using (1) the ''Insertion-deletion (indel) and Carry Forward Correction'' (ICC) pipeline, which clusters sequence reads using a nonsubstitution approach and can correct for indels and carry forward errors, and (2) the ''Primer Identification (ID)'' method, which facilitates construction of a consensus sequence to correct for sequencing errors and allelic skewing. The Primer ID and ICC methods produced similar estimates of viral diversity, but differed in the number of sequence variants generated. Sequence preparation for ICC was comparably simple, but was limited by an inability to assess the number of templates analyzed and allelic skewing. The more costly Primer ID method corrected for allelic skewing and provided the number of viral templates analyzed, which revealed that amplifiable HIV templates varied across specimens and did not correlate with clinical viral load. This latter observation highlights the value of the Primer ID method, which by determining the number of templates amplified, enables more accurate assessment of minority species in the virus population, which may be relevant to prescribing effective antiretroviral therapy.

Human immunodeficiency virus type-1 (HIV) mutates due to the inability of its reverse transcriptase enzyme to proofread.[1–3] The *in vivo* mutation rate is estimated at 3.4×10^{-5} mutations per nucleotide per replication cycle. Thus, given a genome length of $\sim 10^4$ base pairs (bp), and production of 10^{10} new virions each day, the HIV population within an infected individual diversifies rapidly.[1,4]

Drug-resistant variants increase in prevalence under selective drug pressure, but often regress when antiretrovirals are suspended, due to a relatively poor replication capacity. Multiple investigators have sought to determine if low levels of drug-resistant variants in antiretroviral-naive individuals are clinically significant,[5–9] and whether their detection prior to antiretroviral treatment (ART) should alter the choice of antiretrovirals.[8–11] For example, a study of antiretroviral-naive individuals using 454-pyrosequencing found that low levels of mutations conferring resistance to nonnucleoside reverse transcriptase inhibitors (NNRTIs) had a significantly greater risk of virologic failure with an NNRTI-based regimen compared to those without mutations.[12]

To confidently ascribe an attributable risk for virologic failure to minority variants detected by multiple parallel sequencing (MPS) technologies, an adequate number of viral templates must be sampled, with sufficient multiplicity of reads per template. MPS studies of HIV often utilize the clinical plasma viral load as a proxy for the number of templates assayed. However, clinical viral load tests target conserved regions of the HIV genome to minimize the effects of differential primer binding, amplify relatively short spans of the viral template to maximize amplification efficiency, and use internal controls to correct for inefficiencies in extraction and amplification.[13] MPS for drug resistance generally has not utilized similar measures to optimize viral amplification. Consequently, the clinical viral load may overrepresent the

[1]Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, Washington.
[2]Department of Microbiology, University of Washington, Seattle, Washington.
[3]Department of Pediatrics, University of Washington, Seattle, Washington.
[4]Department of Laboratory Medicine, University of Washington, Seattle, Washington.
[5]Department of Medicine, University of Washington, Seattle, Washington.
[6]Department of Global Health, University of Washington, Seattle, Washington.

number of amplifiable templates in ''second-generation deep sequencing technologies,'' including 454-pyrosequencing (Roche), and the sequences generated may be biased by primer sequences, length of amplicons, enzyme fidelity, and PCR conditions.[13,14]

Sequence analysis pipelines can also affect the results of MPS. The Insertion-deletion (indel) and Carry Forward Correction (ICC) pipeline clusters sequence reads using a nonsubstitution approach and corrects for indels and carry forward errors.[15] This pipeline, along with other conventional pipelines, assumes equal amplification of viral templates and therefore that proportions of final reads accurately reflect the diversity of an individual's virus population.

Alternatively, the Primer ID method[16] addresses inaccuracies of estimating template input and corrects skewing of allelic frequencies and PCR and sequencing errors by constructing a consensus sequence for each template that allows for quantification of template input.[16–19] Each viral template is labeled with a unique, randomly generated, 8-bp identifier (i.e., the Primer ID) adjacent to the 5′ end of the gene-specific portion of the primer used to reverse transcribe the viral RNA. The tagging of each template with a unique Primer ID is carried throughout PCR and sequencing, and is used to tabulate the number of viral templates amplified. Three or more Primer ID reads are used to generate a consensus sequence, which corrects PCR and sequencing errors.

In this study the Primer ID method was used to examine the effect of viral load and read depth on the amplified products from HIV *pol* templates. Additionally, the Primer ID method[16] was compared to ICC, to evaluate the correction of sequencing errors and differences in the genetic composition of the final sequence output.

Blood plasma from 12 antiretroviral-naive individuals infected with HIV subtype A from whom sufficient specimen volume was available and plasma HIV RNA was >10K copies/ml was randomly selected for this study and HIV RNA was quantified by a real-time polymerase chain reaction assay (Abbott RealTime HIV-1, Abbott Global). Use of remnant specimens for assay development was approved by the University of Washington's Institutional Review Board.

Nucleic acids extracted using silica[20] from 800 μl of each plasma specimen were resuspended into a volume of 50 μl. Ten microliters of each RNA extract was reverse transcribed (Blueprint 1st Strand cDNA Synthesis Kit, Takara Bio, Inc.) and primers with ID tags (NNNNNNNN) were added as described[16] to reverse transcribe two regions of HIV *pol*: primer 2994R and primer 3267R (Table 1). The 3′ end of both primers anneal to the RT region of HIV *pol* (HXB2 2965 ← 2994, HXB2 3243 ← 3267).[16] Adjacent to the 5′ end of the Primer ID was a region complementary to primer sequences of a subsequent heminested 2nd-round PCR.[16] After cDNA synthesis the solution was purified (High Pure PCR Template Preparation Kit, Roche) to remove excess primers that could potentially tag template in subsequent rounds of PCR.

The primer ID cDNA aliquots were amplified by nested PCR using primers in Table 1 to produce two amplicons of regions encoding HIV reverse transcriptase codons 34–138 (Amplicon 1) and 149–230 (Amplicon 2). Four identical first round reactions, each with 5 μl cDNA, were performed for each sample to avoid inhibition of PCR by adding excess cDNA. Each reaction volume contained cDNA, 0.4 μM of amplicon specific, forward and reverse primers, 2.5 units FastStart High Fidelity Enzyme (Roche), 5 μl 10x FastStart High Fidelity master mix with $MgCl_2$ (Roche), 0.2 mM dNTPs, and dH_2O to equal 50 μl. Five microliters of each first round amplicon was pooled and 2 μl of each pool was amplified in the second round reaction. Second round reagents were identical to that of first round, with 3 μl of dH_2O to bring the reaction volume to 50 μl. A unique 10-base multiplex identifier (MID) was added to each subject's amplicon during a second round PCR to allow sequencing of multiple subjects in one region of the pyrosequencing plate. The first round conditions were 95°C for 2 min, 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min, and a final extension at 72°C for 7 min. The second round conditions were 95°C for 2 min, 30 cycles of 95°C for 30 s, 60°C for 20 s, and 72°C for 45 s, and one final extension at 72°C for 5 min.

The pKS q23-17 DNA plasmid was used as a control[21] for PCR and pyrosequencing errors. The plasmid was amplified with primers 2994R and 2571F in the first round and primers 2621F and ngs2R in the second round (Table 1) to create Amplicon 1, and primers 3267R and 2933F in the first round and primers 2965F and ngs2R in the second round to create Amplicon 2.

Table 1. Primers Used for Amplification of HIV-1 Subtype A

Primer name	*Amplicon 1*	*Position in HXB2*
cDNA synthesis (2994R)	5′-GCCTTGCCAGCACGCTCAGGCCTTGCACGNNNNN NNNTCACATTGTACTGATAYCTRAYWCCTGGTG-3′	2965 ← 2994
1st round forward (2571F)	GTACCAGTAAMATTAAAGCCAGGAATGG	2571 → 2598
1st round reverse (ngs1R)	GCCTTGCCAGCACGCTCAGGC	Not gene specific
2nd round forward (2621F)	MID-GCCATTGACAGAAGARAAAATAAAAGC	2621 → 2647
2nd round reverse (ngs2R)	MID-CCAGCACGCTCAGGCCTTGCA	Not gene specific
Primer name	*Amplicon 2*	*Position in HXB2*
cDNA synthesis (3267R)	5′-GCCTTGCCAGCACGCTCAGGCCTTGCACGNNN NNNNNTCCAYTTGTCAGGATGGAGYTCATA –3′	3243 ← 3267
1st round forward (2932F)	CTGCATTYACYATACCTAGTAYAAAC	2932 → 2957
1st round reverse (ngs1R)	GCCTTGCCAGCACGCTCAGGC	Not gene specific
2nd round forward (2965F)	MID-CACCAGGWRTYAGRTATCAGTACAATGT	2965 → 2992
2nd round reverse (ngs2R)	MID-CCAGCACGCTCAGGCCTTGCA	Not gene specific

After nested PCR, amplicons were purified (High Pure PCR Template Preparation Kit, Roche). The amplicons from each sample were quantified (Quant-iT PicoGreen dsDNA Assay Kit, Invitrogen) and diluted with water to 10^9 copies/μl. An equal volume of primer ID labeled Amplicon 1 and Amplicon 2 fragments from each sample was pooled and 2 μl of the pool was added to 198 μl of nuclease free water to bring the concentration to 10^7 copies/μl for pyrosequencing. One subject's sample (Subject 6) was sequenced in two separate gaskets to assess reproducibility and the impact of doubling the number of reads on the final sequence output.

Primer ID[16] and ICC[15] pipelines were used to analyze the pyrosequencing dataset and generate final sequence variants. The study comparisons included the following: (1) the number of unique IDs generated by the Primer ID method compared to copies input based on viral load; (2) the number of final sequence variants determined by the Primer ID method, performed by condensing monotypic consensus sequences from different primer IDs, vs. the ICC method[15]; (3) the HIV diversity of subjects' final HIV sequence population from each pipeline was compared using the DIVEIN web tool[22]; and (4) the effect of doubling the number of 454-pyrosequencing reads on the number of sequence variants by sequencing the same two amplicons generated from one subject in one vs. two gaskets of the 454-pyrosequencing plate.

A total of 99,696 sequence reads generated by 454-pyrosequencing for the 14 amplicons passed criteria for full length, PCR primer, primer ID, and MID; with two sequenced twice, reads ranged from 4,676 to 11,693 per amplicon. More truncated sequences were observed for Amplicon 1 (317 bp) compared to Amplicon 2 (250 bp) ($p = 0.028$ Wilcoxon sign rank test), most likely due to the greater length of region amplified. While all full-length reads were analyzed using the ICC pipeline, only those with unique IDs ≥3 reads were analyzed using the Primer ID method, which variably affected the number of reads by amplicon and specimen. The number of reads per each unique primer ID was not uniform, but progressively diminished, as shown for Subject 4 (Fig. 1). Across subjects and amplicons, 7% of reads were discarded due to incomplete length and 61% due to <3 reads per Primer ID (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/aid).

The input number of HIV templates and the template length/region amplified affected the yield of 454-prosequencing reads using the Primer ID method. The yield of unique IDs (any number of reads) was substantially less than the estimated input based on clinical viral load across all samples (data not shown), as was the number of unique IDs ≥3 (Fig. 2A). Relatively few unique IDs ≥3 were observed when a low or a high input of viral templates, as estimated by clinical viral load, was assayed. The proportion of viral templates with Primer IDs ≥3 sequences generated by a set number of 454-pyrosequencing reads diminished with increasing input of viral templates, based on the clinical viral load assay (Fig. 2B). Increasing the number of reads for a subject's specimen, shown by combining the sequence data from assaying Subject 6's amplicons in two different 454-plates, found that the number of raw reads increased in an additive fashion and the number of unique IDs with three or more reads also increased (Table 2). After combining the Primer ID consensus sequences with identical HIV *pol* sequences the number of final sequence variants increased for Amplicon 1 but results for Amplicon 2 were varied (Table 2).

To reduce sequencing expenses, some researchers may choose to forgo the Primer ID method. A comparison of the number of final sequence variants generated by Primer ID to ICC using a Wilcoxon sign rank test at a 95% confidence level found significantly more variants by Primer ID for Amplicon 1 ($p = 0.028$) and no significant difference for Amplicon 2 ($p = 0.25$).

To compare the genetic makeup of the populations generated by each method, phylogenetic analyses were performed for each subject and diversity was calculated. Using a Wilcoxon sign rank test at 95% confidence, the diversities produced by each method were not significantly different ($p = 0.173$ for Amplicon 1 and $p = 0.25$ for Amplicon 2).

Our comparison of the number of HIV templates sequenced across two regions of *pol* from specimens with a range of clinical viral loads demonstrates several pitfalls with MPS that can lead to insensitive assays for minority species of the HIV population. The Primer ID method in our study produced significantly fewer unique sequences than estimated by the clinical viral load, only <1% to 14% of the input estimated by viral load, which overlaps with the range

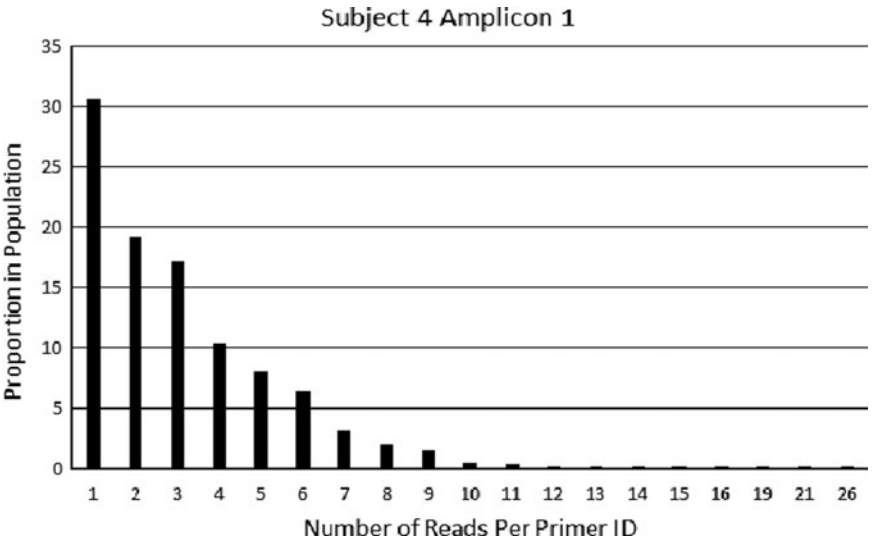

FIG. 1. The distribution of the number of sequence reads per unique Primer Identification tags (IDs) is shown for Subject 4's Amplicon 1. The reads shown cover the entire amplicon, and passed criteria for PCR primer, primer ID, and Multiplex Identifier (MID). The distribution was similar for Amplicon 2 from this subject, but other subjects had an even larger proportion of reads discarded due to a larger proportion of 1 or 2 reads per ID. Input for this specimen was 11,355 copies based on Abbott RealTime HIV-1 assay.

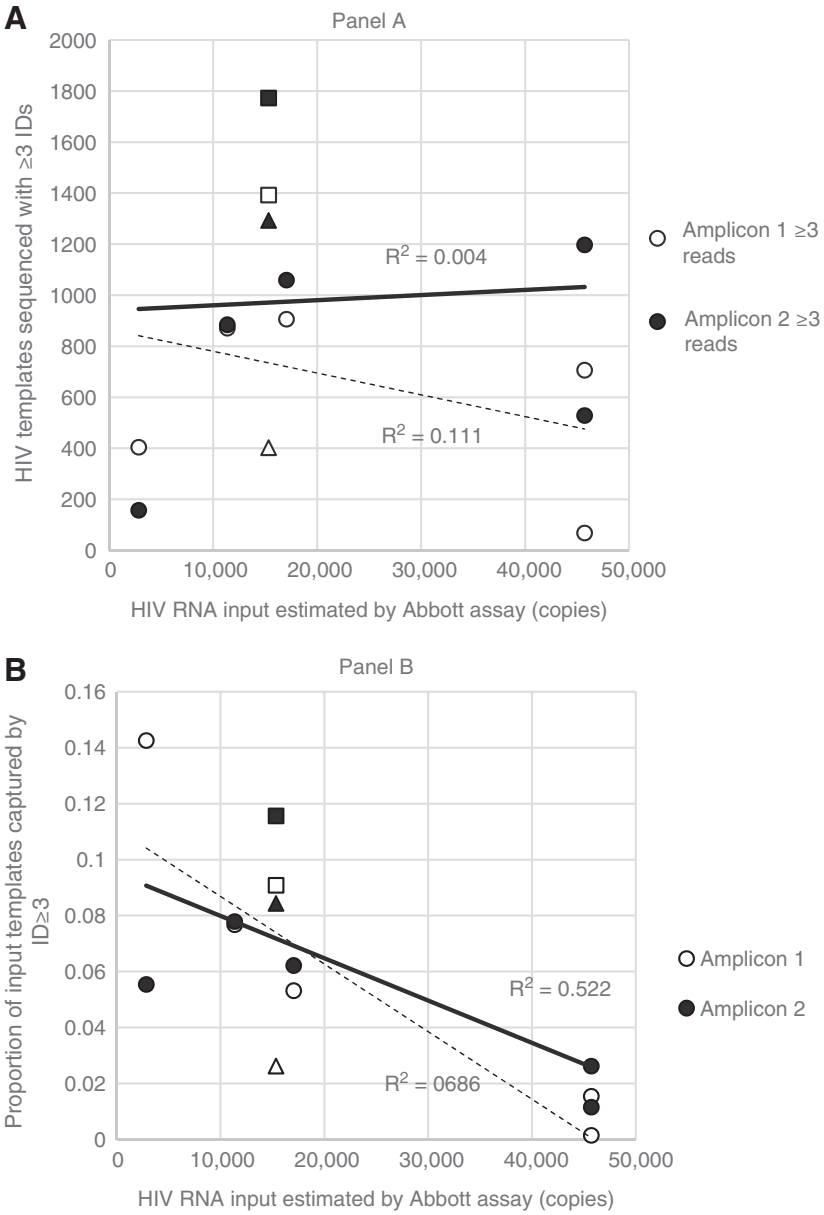

FIG. 2. The yield of human immunodeficiency virus (HIV) sequences labeled with unique Primer Identification tags (IDs) is affected by input of viral templates and number of sequence reads. **(A)** The number of HIV templates sequenced after filtering for PCR primer, primer ID, Multiplex Identifier (MID), and sequence length relative to the input of HIV RNA as quantified by clinical viral load assay. Trends are compared for Amplicon 1 ≥ 3 reads and Amplicon 2 ≥ 3 reads (all IDs meeting filter and having ≥3 reads per ID). **(B)** The proportion of input HIV RNA templates captured by using ≥3 reads as criteria for inclusion in the final data set. **(A, B)** A *triangle* is used to denote the average of Subject 6's read from plate-1 and plate-2, while a s*quare* is used to denote Subject 6's results when reads from both plates were combined. The *dashed trend line* corresponds to Amplicon 1 while the *solid trend line* corresponds to Amplicon 2.

reported by others.[23] Submission of too many viral templates in our study reduced the proportion of templates sequenced. Alternatively, submission of too few templates would not allow for detection of minority variants. Furthermore, the proportion of templates amplified from a specific specimen varied and did not correlate directly with clinical viral load. Factors that may diminish amplification of templates include specimen integrity, length of amplicon, properties of reverse transcriptases and DNA polymerases, and primer binding to specific templates.

Table 2. Effect of Doubling the Number of Sequence Reads on the Number of HIV Final Sequence Variants for Subject #6

	Amplicon 1[a]			*Amplicon 2*[a]		
Subject #	*Reads pass filter*[b]	*Unique primer ID ≥3 reads*[c]	*Final sequence variants*[d]	*Reads pass filter*[b]	*Unique primer ID ≥3 reads*[c]	*Final sequence variants*[d]
6 (plate-1)	6,063	450	223	8,826	1,184	95
6 (plate-2)	5,418	354	344	11,693	1,402	640
6 (combined)	11,481	1,393	906	20,519	1,773	386

[a]Subject 6's amplicon was sequenced twice, in two pyrosequencing plates.
[b]The number of sequence reads that passed the filter (PCR primer, primer ID, MID, covering the full amplicon).
[c]Unique Primer IDs with ≥3 reads.
[d]Number of final sequence variants generated from combining identical HIV sequences.

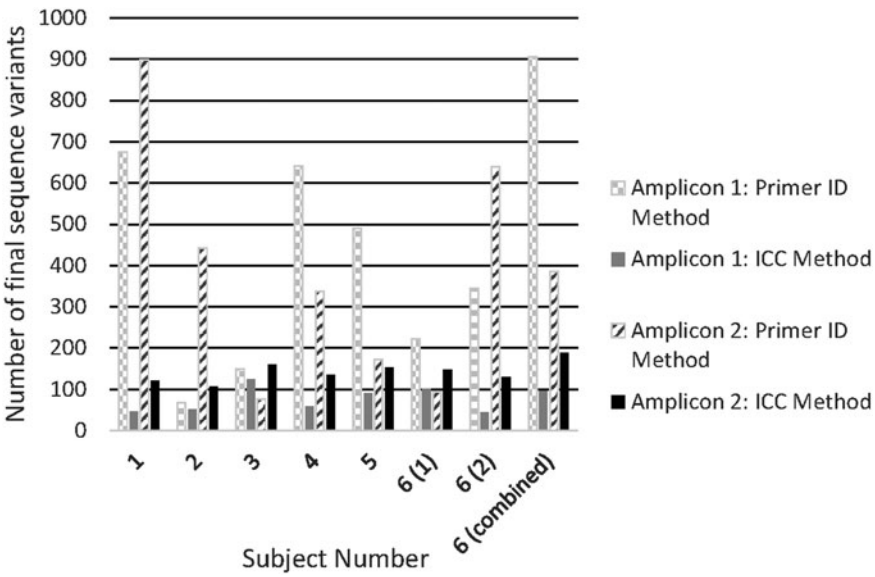

FIG. 3. Number of final sequence variants analyzed after processing by either Primer Identification (ID) (using ≥3 reads for each consensus sequence) or Insertion-deletion (indel) and Carry Forward Correction (ICC) pipelines showing Amplicon 1 and Amplicon 2 separately. The Primer ID method produced more final sequence variants than the ICC method for Amplicon 1, but not for Amplicon 2 ($p = 0.028$ and $p = 0.25$, respectively, using the Wilcoxon sign rank test at 95% confidence level), most likely because the shorter length of the template amplified produced a greater number of Primer ID reads for Amplicon 2, which allowed for more unique IDs with ≥3 reads.

Sequence analysis ''pipelines'' aim to increase the accuracy of data, however, elimination of poor-quality data further reduces the yield of sequences. The fraction of non-full-length reads generated by 454-pyrosequencing increases with amplicon length, as shown by our yield from Amplicon 1 compared to Amplicon 2 (Supplementary Table S1). In addition, to correct for misinsertion of bases during PCR, the Primer ID method discards sequence reads with Primer IDs not represented by three or more reads. By increasing the number of sequence reads for a given sample, as accomplished by assaying two aliquots of Subject 6's amplicon, the unique Primer IDs with ≥3 reads increased in number. Clearly, relatively few reads per template limited the proportion of viral templates sequenced across several of the specimens we assayed, and emphasizes the need to optimize the multiplicity of reads per viral templates amplified. Additionally, by increasing the reads per template for Subject 6, a greater proportion of our final sequences appears to be true variants due to better error correction.

Analysis of 454-pyrosequencing without use of Primer IDs and processed using ICC was not found to yield significantly different numbers of sequence variants for Amplicon 2, but the two methods differed for Amplicon 1, with the ICC method producing fewer final sequence variants for Amplicon 1 (Fig. 3). This difference might have occurred due to diverse handling of errors inherent in 454-pyrosequencing. Analysis by Primer ID combines sequences into a consensus by identity of the 8-bp random sequences added at the time of reverse transcription, whereas ICC groups identical sequences across the length of HIV *pol* into one final sequence variant. In ICC, if two sequences were not grouped together due to error (mismatch, indel, carry forward) this would generate a larger number of final sequence variants. Increasing the number of reads per template should make more Primer IDs available for analysis, which should increase the number of templates with ≥3 reads and reduce all the aforementioned types of error by combining variants into one consensus sequence; this may have occurred with Amplicon 2 of Subject 6 (Fig. 3) where

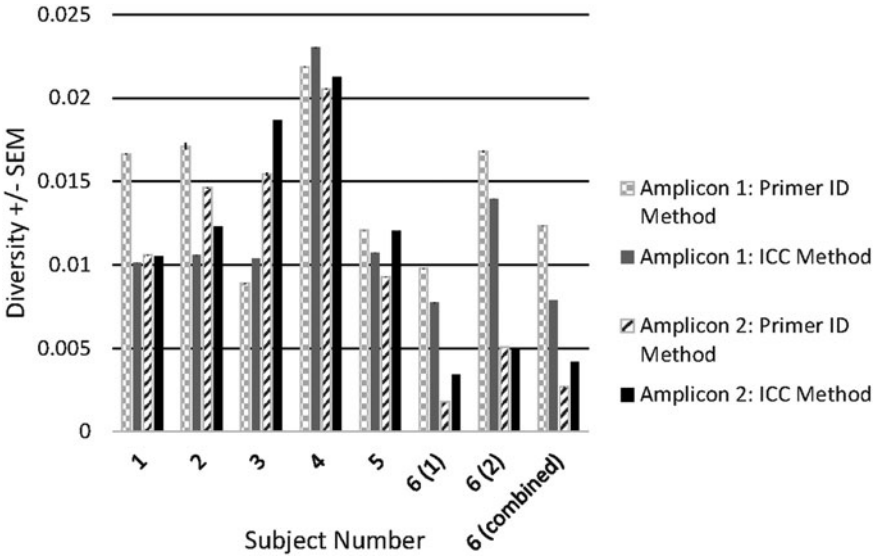

FIG. 4. The pairwise diversity of human immunodeficiency virus (HIV) *pol* calculated using sequences generated by the Primer Identification (ID) and the Insertion-deletion (indel) and Carry Forward Correction (ICC) pipelines, with standard errors (bars are very small). The diversity calculated for each amplicon appears similar between the two methods ($p = 0.173$ for Amplicon 1 and $p = 0.25$ for Amplicon 2 using the Wilcoxon sign rank test at 95% confidence level).

increasing the number of reads decreased the final sequence variants. The addition of more sequences by combining the pools for Subject 6 appears to have condensed some previously unique consensus sequences into one final sequence variant, which has improved error correction.

The increase in final sequence variants for Subject 6's Amplicon 1 demonstrates that with more reads more novel sequences were available since fewer IDs were discarded due to the requirement of three or more reads, thus more IDs with greater than three reads than the sum of plates 1 and 2 are included in the analysis. However, fewer Primer IDs had the same consensus sequence, thus fewer were combined for the final set of sequence variants because a sequence variant is defined as a unique sequence, not a unique ID, resulting in more final sequence variants than the original sum of plates 1 and 2. This could be because Amplicon 1 covers a naturally more diverse region in Subject 6's virus and by increasing reads more unique viral variants were sequenced. Alternatively, the longer length of Amplicon 1 or increased homopolymers in Amplicon 1 (27 homopolymers) compared to Amplicon 2 (18 homopolymers) could artificially inflate the number of final sequence variants. In contrast, the ICC method corrects indel and carry forward errors at the sequence level by nonsubstitution sequence clustering,[15] and subsequently applies a statistical model to identify ''true'' variants using a known sequencing error distribution. This error correction allows ICC to group sequences into a final consensus sequence, allowing a higher template to read ratio for error correction. It is also possible that the ICC method overcorrects, and in doing so produces fewer final sequence variants. Despite these differences in analysis, the viral diversity by the two methods were similar, with no significant differences or trends between the two methods (Fig. 4).

Importantly, ICC does not provide the number of viral templates evaluated in the assay, which is provided by the Primer ID method, and is critical for assessing minority variants in a viral population. However, the quantification of viral templates by the Primer ID method could be artificially increased due to substitution errors in the 8-bp Primer ID.[24,25] Additionally, as the use of only eight random base combinations for the Primer ID generates 4^8 or 65,536 different IDs, input of a large number of viral templates will increase the probability that each unique 8-bp sequence labels more than one template, which would underestimate the number of templates evaluated.[25,26]

The reports that HIV minority variants appear clinically relevant and lead to virologic failure of ART[5,6, 8–11] emphasize the importance of data representative of the true proportions of variants in the viral population. Administration of ART that inhibits all replication-competent variants in an individual can reduce virologic failure and minimize the spread of HIV, especially the spread of virus that is already resistant to certain medications. To provide the most efficacious ART for each patient, accurate testing of their viral population appears crucial. Adequate sampling of the viral population is implicit to quantifying minority variants, as is designing primers likely to amplify all variants. Our data show that the use of the clinical viral load does not accurately determine amplifiable viral templates, and that amplification varies across specimens, due to amplicon length, enzyme fidelity, PCR conditions, presumed primer biases, and/or inhibitory factors in the specimen.[13,14] Omission of PCR by direct sequencing of sheared nucleic acids using the Illumina platform eliminates primer biases, but adds cost, as this method generates comparably more human to viral sequences. The Illumina platform offers a lower frequency of indel errors, and the more recent generations of the assay have read lengths comparable to the 454-pyrosequencing platform.

The sequencing of relatively few HIV templates compared to the predicted input from clinical viral loads as well as the variability of the number of templates sequenced by gene region and individual's specimen all substantiate the utility of the Primer ID method or direct sequencing of sheared nucleic acids to quantify the number of viral templates sequenced to quantify minority variants. Further comparative studies are required to determine the relative advantages and costs of these two methods as well as other methods.

Acknowledgments

We appreciate and thank the individuals who volunteered for this study and the technical contributions made by Rachel Payant and Sheila Strychak.

This work was supported by NIH funds through an International Maternal Pediatric Adolescent AIDS Clinical Trials Group (IMPAACT) Virology Developmental Laboratory award (L.M..F) (U01 AI068632) and the Clinical Research and Retrovirology Core of the Seattle Centers for AIDS Research (P30 AI027757) (King Holmes). Work was also supported by R21 AI084688 to L.M.F. Overall support for IMPAACT was provided by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under award numbers UM1AI068632 (IMPAACT LOC), UM1AI068616 (IMPAACT SDMC), and UM1AI106716 (IMPAACT LC), with co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and the National Institute of Mental Health (NIMH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Disclosure Statement

No competing financial interests exist.

References

1. Taylor BS, Sobieszczyk ME, McCutchan FE, and Hammer SM: The challenge of HIV-1 subtype diversity. N Engl J Med 2008;358(15):1590–1602. Erratum: N Engl J Med 2008;359(18):1965–1966, 1972.
2. Roberts JD, Bebenek K, and Kunkel TA: The accuracy of reverse transcriptase from HIV-1. Science 1988;242(4882): 1171–1173.
3. Preston BD, Poiesz BJ, and Loeb LA: Fidelity of HIV-1 reverse transcriptase. Science 1988;242(4882):1168–1171.
4. Mansky LM and Temin HM: Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 1995;69(8):5087–5094.
5. Clavel F and Hance AJ: HIV drug resistance. N Engl J Med 2004;350(10):1023–1035.
6. Li JZ and Kuritzkes DR: Clinical implications of HIV-1 minority variants. Clin Infect Dis 2013;56(11):1667–1674.
7. Varghese V, Shahriar R, Rhee SY, *et al.:* Minority variants associated with transmitted and acquired HIV-1 nonnucleoside

reverse transcriptase inhibitor resistance: Implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. J Acquir Immune Defic Syndr 2009;52(3):309–315.
8. Johnson JA, Li JF, Wei X, *et al.:* Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. PLoS Med 2008;5(7):e158.
9. Metzner KJ, Giulieri SG, Knoepfel SA, *et al.:* Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and -adherent patients. Clin Infect Dis 2009;48(2):239–247.
10. Li JZ, Paredes R, Ribaudo HJ, *et al.:* Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: A systematic review and pooled analysis. JAMA 2011;305(13):1327–1335.
11. Halvas EK, Wiegand A, Boltz VF, *et al.:* Low frequency nonnucleoside reverse-transcriptase inhibitor-resistant variants contribute to failure of efavirenz-containing regimens in treatment- experienced patients. J Infect Dis 2010;201(5): 672–680.
12. Simen BB, Simons JF, Hullsiek KH, *et al.:* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 2009;199(5):693–701.
13. Larsen BB, Chen L, Maust BS, *et al.:* Improved detection of rare HIV-1 variants using 454 pyrosequencing. PLoS One 2013;8(10):e76502.
14. Mallona I, Weiss J, and Marcos E-C: pcrEfficiency: A Web tool for PCR amplification efficiency prediction. BMC Bioinformatics 2011;12:404.
15. Deng W, Maust BS, Westfall DH, *et al.:* Indel and Carryforward Correction (ICC): A new analysis approach for processing 454 pyrosequencing data. Bioinformatics 2013;29(19):2402–2409.
16. Jabara CB, Jones CD, Roach J, *et al.:* Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA 2011;108(50):20166–20171.
17. Kanagawa T: Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng 2003;96(4): 317–323.
18. Suzuki MT and Giovannoni SJ: Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microbiol 1996;62(2):625–230.
19. Polz MF and Cavanaugh CM: Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol 1998;64(10):3724–3730.
20. Boom R, Sol CJ, Salimans MM, *et al.:* Rapid and simple method for purification of nucleic acids. J Clin Microbiol 1990;28(3):495–503.
21. Poss M and Overbaugh J: Variants from the diverse virus population identified at seroconversion of a clade A human immunodeficiency virus type 1-infected woman have distinct biological properties. J Virol 1999;73(7):5255–5264.
22. Deng W, Maust BS, Nickle DC, *et al.:* DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and informative sites. Biotechniques 2010;48(5):405–408.
23. Keys JR, Zhou S, Anderson JA, *et al.:* Primer ID informs next-generation sequencing platforms and reveals preexisting drug resistance mutations in the HIV-1 reverse transcriptase coding domain. AIDS Res Hum Retroviruses 2015;31:658–668.
24. Zhou S, Jones C, Mieczkowski P, and Swanstrom R: Primer ID validates template sampling depth and greatly reduces the error rate of next-generation sequencing of HIV-1 genomic RNA populations. J Virol 2015;89(16):8540–8555.
25. Liang RH, Mo T, Dong W, *et al.:* Theoretical and experimental assessment of degenerate primer tagging in ultradeep applications of next-generation sequencing. Nucleic Acids Res 2014;42(12):e98.
26. Sheward DJ, Murrell B, and Williamson C: Degenerate primer IDs and the birthday problem. Proc Natl Acad Sci USA 2012;109(21):E1330; author reply E1331.

Address correspondence to:
Lisa Frenkel
1900 9th Avenue
Seattle, Washington 98101

E-mail: lfrenkel@uw.edu