AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

VIROLOGY

A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings

David Bonsall,[a,b] Tanya Golubchik,[a] Mariateresa de Cesare,[a,b] Mohammed Limbada,[c,d] Barry Kosloff,[c,d] George MacIntyre-Cockett,[b,a] Matthew Hall,[a] Chris Wymant,[a] M. Azim Ansari,[b,e] Lucie Abeler-Dörner,[a] Ab Schaap,[c,d] Anthony Brown,[e] Eleanor Barnes,[e] Estelle Piwowar-Manning,[f] Susan Eshleman,[f] Ethan Wilson,[g] Lynda Emel,[g] Richard Hayes,[d] Sarah Fidler,[h] Helen Ayles,[c,d] Rory Bowden,[b] Christophe Fraser,[a,b] HPTN 071 (PopART) Team

[a]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[b]Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
[c]ZAMBART, University of Zambia, Lusaka, Zambia
[d]London School of Hygiene and Tropical Medicine, London, United Kingdom
[e]Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom
[f]Dept. of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[g]Statistical Centre for HIV/AIDS Research, Fred Hutchinson Cancer Research Centre, Seattle, Washington, USA
[h]Department of Infectious Disease, Imperial College London, Imperial College NIHR BRC, London, United Kingdom

David Bonsall and Tanya Golubchik contributed equally to this work, and the order in which they are listed reflects a decreasing order of managerial responsibility for the project.

ABSTRACT Viral genetic sequencing can be used to monitor the spread of HIV drug resistance, identify appropriate antiretroviral regimes, and characterize transmission dynamics. Despite decreasing costs, next-generation sequencing (NGS) is still prohibitively costly for routine use in generalized HIV epidemics in low- and middle-income countries. Here, we present veSEQ-HIV, a high-throughput, cost-effective NGS sequencing method and computational pipeline tailored specifically to HIV, which can be performed using leftover blood drawn for routine CD4 cell count testing. This method overcomes several major technical challenges that have prevented HIV sequencing from being used routinely in public health efforts; it is fast, robust, and cost-efficient, and generates full genomic sequences of diverse strains of HIV without bias. The complete veSEQ-HIV pipeline provides viral load estimates and quantitative summaries of drug resistance mutations; it also exploits information on within-host viral diversity to construct directed transmission networks. We evaluated the method's performance using 1,620 plasma samples collected from individuals attending 10 large urban clinics in Zambia as part of the HPTN 071-2 study (PopART Phylogenetics). Whole HIV genomes were recovered from 91% of samples with a viral load of >1,000 copies/ml. The cost of the assay (30 GBP per sample) compares favorably with existing VL and HIV genotyping tests, proving an affordable option for combining HIV clinical monitoring with molecular epidemiology and drug resistance surveillance in low-income settings.

KEYWORDS HIV, NGS, viral genomics, public health, sub-Saharan Africa, viral sequencing, bait capture, short-read sequencing, Illumina, SMARTer, HPTN, PopART, HPTN 071, phylogenetics, viral evolution, drug resistance, antiretroviral therapy, RNA virus, antiretroviral resistance, drug resistance evolution, gene sequencing, human immunodeficiency virus, phylogenetic analysis, surveillance studies

Achieving sustained reductions in the incidence of HIV infections through programs of universal access to testing and antiretroviral treatment (UTT) remains a major goal in public health. International efforts have been focused on working toward the UNAIDS "90-90-90" targets, with 90% of people living with HIV (PLWH) diagnosed, 90%

Citation Bonsall D, Golubchik T, de Cesare M, Limbada M, Kosloff B, MacIntyre-Cockett G, Hall M, Wymant C, Ansari MA, Abeler-Dörner L, Schaap A, Brown A, Barnes E, Piwowar-Manning E, Eshleman S, Wilson E, Emel L, Hayes R, Fidler S, Ayles H, Bowden R, Fraser C, HPTN 071 (PopART) Team. 2020. A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings. J Clin Microbiol 58:e00382-20. https://doi.org/10.1128/JCM.00382-20.

Editor Angela M. Caliendo, Rhode Island Hospital

Address correspondence to David Bonsall, david.bonsall@bdi.ox.ac.uk.

Received 2 March 2020
Returned for modification 28 March 2020
Accepted 10 July 2020
Accepted manuscript posted online 15 July 2020
Published 22 September 2020

of those on antiretroviral therapy (ART), and 90% of those successfully virally suppressed (1, 2). HIV drug resistance compromises the ability of ART to suppress viral replication. The frequency of drug resistance is expected to increase as UTT becomes more common (3), which may make it difficult to reach the WHO goal. A 2017 report by the WHO identified parts of the world where more than 10% of people living with HIV already harbor virus resistant to current first line antiretroviral drugs (4). This has driven the switch to dolutegravir-based regimens as preferred first line ART.

Both the spread of drug resistance and transmission patterns can be better understood by analyzing viral sequence data (5). To date, clinical drug resistance testing has primarily relied on Sanger consensus sequencing of HIV *pol* genes. Next-generation sequencing (NGS) also produces detailed minority variant information, which can detect low-frequency drug resistant viral variants. However, despite its benefits, adoption of NGS for HIV drug resistance testing has been slow, in part due to technical difficulties in obtaining whole-genome sequences for all genotypes, particularly at low viral loads, and uncertainty over distinguishing low-frequency mutations from the sequencing artifacts and contamination that occur during massive parallel sequencing. Recently, the FDA approved the first NGS assay for HIV drug resistance using *pol*-specific PCR that can sequence up to 15 samples in parallel (6); however, demand remains for more high-throughput, low-cost options for use clinically and as a surveillance tool in high-prevalence settings. In addition, within an appropriate ethical framework, NGS enhances resolution in transmission analyses, indicating transmission direction and thus revealing population characteristics of transmitters and recipients (7). The potential for viral whole-genome sequencing to transform global health surveillance operations has been noted (8).

Large-scale NGS sequencing of HIV genomes using virus-specific PCR (9) has been used to produce whole viral genomes for European samples (10), but the method's performance was found to be far from optimal for analysis of sub-Saharan African samples, with amplification failures resulting in biased genome coverage (11). We previously described veSEQ, a probe-based enrichment method, free of virus-specific PCR, which can be used to sequence viruses directly from clinical samples (12). Here, we describe veSEQ-HIV, a comprehensive laboratory and computational protocol specifically developed to support clinical management and public health programs in low-income settings.

MATERIALS AND METHODS

Samples. Patients were recruited to the HPTN 071-2 (PopART phylogenetics) study by research assistants at 10 urban primary health care facilities, located in 9 of the 12 Zambian communities of the main trial (one community had two health care facilities) (13). The nine communities involved were evenly split between the three study arms of HPTN-071. Patients were recruited if they were aged 18 or over, not currently taking ART, and if they specifically consented to the ancillary phylogenetic study. Most patients were either newly enrolled in the clinic or enrolled and newly eligible for ART; a small fraction was recruited having recently missed several doses of ART. The study protocol (https://www.hptn.org/sites/default/files/inline-files/HPTN%20071-2%2C%20Version%202.0%20%2807-14-2017%29.pdf) has been approved by the ethics committees of the University of Zambia (c/o the Zambian ministry of health) and of the London School of Hygiene and Tropical Medicine.

Sampling. No additional blood samples were required for this ancillary study. Unused samples of blood collected from consenting individuals undergoing routine CD4 cell count testing were transported to the local hospital on the same day. Blood was centrifuged twice and two 500-μl aliquots of plasma were frozen at −80°C. Samples were transported to a central research laboratory (ZAMBART facility) in Lusaka, Zambia using a mobile −20°C freezer, and then shipped to the sequencing laboratory in the United Kingdom. Samples were processed approximately in order of collection and represented the diversity of the population recruited at the beginning of the study.

Laboratory methods. Total RNA was extracted with magnetized silica from HIV-infected plasma lysed with guanidine thiocyanate and with ethanol washes and elution steps performed using the NUCLISENS easyMAG system (bioMérieux). The total 30 μl elution volume was reduced with Agencourt RNAClean XP (Beckman Coulter) to maximize the input RNA mass while minimizing volume for library preparation.

Libraries retaining directionality were prepared using the SMARTer Stranded Total RNA-Seq kit v2 - Pico Input Mammalian (Clontech, TaKaRa Bio) with the following protocol specifications. Total RNA was first denatured at 72°C with the addition of tagged random hexamers to prime reverse transcription, followed by cDNA synthesis according to the manufacturer's protocol option with no fragmentation. The

first strand cDNA was then converted into double-stranded dual-indexed amplified cDNA libraries using in-house sets of 96 i7 and 96 i5 indexed primers (14), using a maximum of 12 PCR cycles. All reaction volumes were reduced to one quarter of the SMARTer kit recommendation and set up was either prepared manually or automated using the Echo 525 (Labcyte) low-volume liquid handler.

No depletion of ribosomal cDNA was carried out prior to target enrichment. Equal volumes (5 μl from a total of 12.5 μl) of each amplified library were pooled in 96-plex without prior cleanup. The pool was cleaned with a lower ratio of Agencourt AMPure XP than recommended by the SMARTer protocol, to eliminate shorter libraries (0.68×). The size distribution and concentration of the 96-plex was assessed using a High Sensitivity D1000 ScreenTape assay on a TapeStation system (Agilent) and a Qubit dsDNA HS Assay (Thermo Fisher Scientific).

A total of 500 ng of pooled libraries was hybridized (SeqCap EZ reagent kit, Roche) to a mixture of custom HIV-specific biotinylated 120-mer oligonucleotides (xGen Lockdown Probes, Integrated DNA Technologies), then pulled down with streptavidin-conjugated beads as previously reported (12). Unbound DNA was washed off the beads (SeqCap EZ hybridization and wash kit, Roche), and the captured libraries were then PCR amplified to produce the final pool for sequencing using a MiSeq (Illumina) instrument with v3 chemistry for a read length up to 300 nt paired-end. Alternatively, up to 384 samples were sequenced on HiSeq 2500 set to Rapid run mode using HiSeq Rapid SBS kit v2 with maximum read lengths of 250 nt.

To confirm assay quantivity, clinical viral load measurements were obtained for 146 specimens also sequenced with veSEQ-HIV. Oxford University Hospital's clinical microbiology laboratory used the COBAS AmpliPrep/COBAS TaqMan HIV1 Test (Roche Molecular Systems, Branchburg, NJ, USA).

Computational pipeline. Raw sequencing reads were first processed with Kraken (15) to identify human and bacterial reads. Kraken was run with default parameters (k = 31 with no filtering), using a custom database containing the human genome together with all bacterial, archaeal and viral genomes from RefSeq, a subset of fungal genomes, and all 9,049 complete HIV genomes from GenBank (last updated 18 May 2018). Reads were filtered to retain only viral and unclassified sequences, and these were trimmed to remove adaptors and low-quality bases using Trimmomatic (16), retaining reads of at least 80 bp. Filtered, trimmed sequences were assembled into contigs using SPAdes (17) and metaSPAdes (18) with default parameters for k (21 to 127). Contiguous sequences assembled from both assembly runs were clustered using cd-hit-est to remove redundant contigs (19), retaining the longest sequence in each cluster with a minimum sequence identity threshold of 0.9. Contigs, together with the filtered reads, were then used to generate HIV genomes and variant frequencies using *shiver* (20), with position-based deduplication of reads enabled. Samples for which no contigs could be assembled were mapped to the closest known HIV reference as identified by Kallisto (21), hashing the filtered reads against a set of 199 HIV reference genomes from the Los Alamos HIV database (http://www.hiv.lanl.gov/), and taking the closest matching genome as the mapping reference for *shiver*. veSEQ-HIV is quantitative, in that the total amount of sequences recovered correlated with viral load. This arises because PCR conditions remain nonsaturating and unbiased probes are used for virus enrichment. A further slight improvement is obtained by computationally removing duplicate copies of viral fragments from sequence data, which are generated by non-virus-specific PCR steps in the protocol. The sequence-derived viral load, in copies/ml, was calculated from the number of deduplicated HIV reads for each sample, using a linear regression model derived from a subset of 146 samples for which we obtained an independent, clinically measured viral load. The R^2 value for this model was 0.89, with no evidence of bias and a mean squared error of prediction of 0.324 $\log_{10}$ copies/ml. The model was used to estimate a sequence-derived viral load for the full data set.

A panel of quantification standards was used to ensure quantitativity and guard against batch effects. The standards comprised five dilutions of subtype B virus spiked into plasma (AcroMetrix HIV-1 Panel copies/ml, Thermo Fisher Scientific), and either one or two negative plasma controls. These were grouped with each batch of 90 HPTN 071-2 (PopART phylogenetics) samples at the point of RNA extraction. We first introduced these standards in batch 6, and have been using these to monitor the quantitativity of each batch.

Contaminant reads were identified and removed using *phyloscanner* for in-depth analyses of *pol* sequencing data. *Phyloscanner* contains several procedures not only for detecting contaminant reads in NGS data sets (7), but also for "blacklisting" them (specifically removing them from consideration for further analysis). Blacklisting works by identifying reads in a sample that are either (i) identical to those from a second sample but present in much smaller numbers, or (ii) are phylogenetically distant from the majority of the sample's reads and are relatively few in number. A total of 373 overlapping genomic windows, each of length 340 bp, were selected, staggering the starting positions by 5 bp. For each 340-bp window, a phylogeny was inferred for all read pairs that fully spanned that window, and ancestral state reconstruction divided the reads for each sample into distinct groups (subgraphs), with the *phyloscanner* Sankoff k parameter set to 12.5. A group of reads was flagged as likely contamination if it contained three or fewer reads, or less than 0.1% of the total number of reads from the sample in that window. The consensus sequence and minority base frequencies were then recalculated from the resulting cleaned mapped reads using *shiver* (20). The complete workflow is included within *phyloscanner* ("phyloscanner clean"). "Phyloscanner clean" can be further optimized where the approximate proportion of expected contaminant reads is known, e.g., from laboratory controls.

Finally, both the consensus sequence and the cleaned reads were analyzed with the Stanford drug resistance tool (22) to determine consensus and minority drug resistance levels. Aggregated drug resistance predictions, accounting for mutations linked on the same read pair, were calculated as the

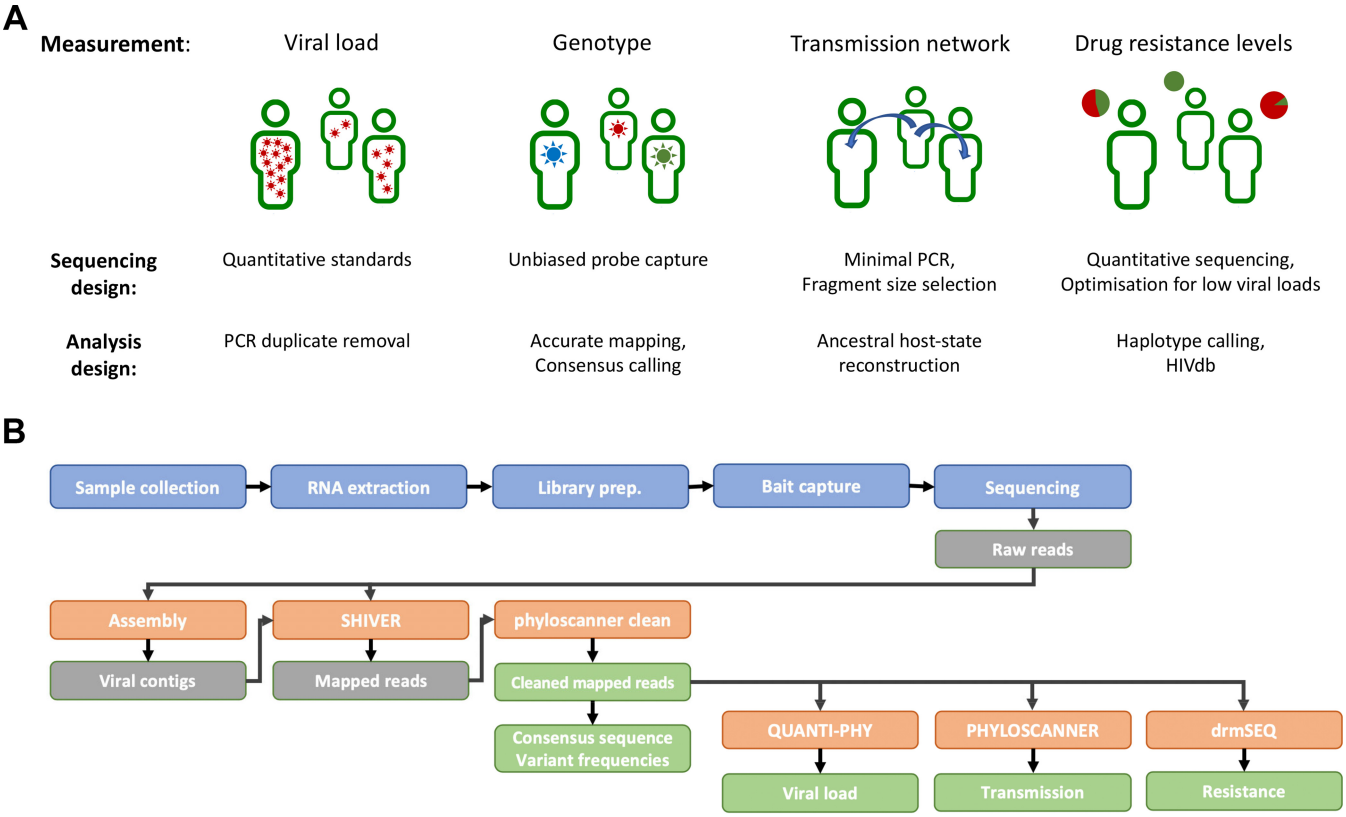

FIG 1 veSEQ-HIV includes a sequencing protocol and bioinformatics pipeline, yielding information on individual and population levels. (A) The veSEQ-HIV method was developed to provide multiple measurements from a single assay, including viral load, HIV genotype, drug resistance, and transmission inference. (B) Overview of veSEQ-HIV: a complete laboratory and computational pipeline for high-throughput sequencing. RNA extraction from plasma samples is carried out in a CL-3 certified laboratory, before transfer to a dedicated genomics facility for library preparation, bait capture, and finally sequencing. Raw sequencing data are preprocessed to remove host and contaminant RNA, and these computationally filtered reads together with their *de novo*-assembled contigs are used to determine the consensus genome and minority variant frequencies using *shiver*. QC metrics are then calculated, and the proportion of contaminant reads originating from other samples is estimated with Kallisto. Samples which result in a successful read mapping are then cleaned with *phyloscanner* to remove contaminant reads, and clean reads are used to infer transmission patterns with *phyloscanner*, and to make drug resistance predictions with HIVdb and drmSEQ.

maximum level of resistance (susceptible < potential low-level < low-level < intermediate < high-level) observed in at least 20% of merged read pairs spanning each position.

RESULTS

The veSEQ-HIV protocol was developed to obtain multiple measurements from a single assay (Fig. 1). It provides a quantitative viral load estimate across at least 5 orders of magnitude, frequency of drug resistance mutations at both consensus and minority variant levels, and accurate and unbiased genotype information that is suitable for ancestral state reconstruction and the generation of directed transmission networks.

The method is the integration of a laboratory protocol with a bioinformatics pipeline (Fig. S1 in the supplemental material). Briefly, RNA extraction is followed by library preparation, bait capture, and sequencing. The bioinformatics pipeline removes host and contaminant RNA, and constructs consensus genome and minority variant frequencies from *de novo* assembled contigs using *shiver* (20). *Phyloscanner* is used to remove contaminant reads and infer transmission patterns (7). Drug resistance predictions are made with reference to HIVdb (Stanford database) (22).

veSEQ-HIV is robust and cost-effective. The development of veSEQ-HIV was achieved by optimizing veSEQ, our sequencing method for hepatitis C virus (HCV) (12). Our aims were to increase sensitivity and throughput, while minimizing cost, processing time, and protocol complexity. Compared to the enzymatic method for adapter ligation used in the original veSEQ protocol, the SMARTer protocol (Switching mechanism at 5′ end of RNA template) produced more unique (PCR deduplicated) sequences

per sample, required fewer protocol steps and disposable plastics, and required no pre-PCR buffer exchanges (23). By concentrating extracted nucleic acids, (RNAClean XP) SMARTer reagent volumes could be reduced 4-fold without loss in library complexity. Automation was achieved in 96- and 384-well formats using 96-channel pipettes (PlateMaster, Gilson). The steps of the final protocol are listed in Table S1.

Like most high-throughput NGS protocols, veSEQ-HIV requires fragmentation of the virus RNA into so-called "inserts." In previous work, we found that inserts of 350 bp or more offer useful insights into within-host phylogenetic diversity (7); we therefore sought to optimize the length of these inserts to be as long as possible within the limits compatible with the Illumina sequencing platform (350 to 600bp). After reducing preenrichment PCR cycles from 18 to 12 and introducing a size-selective bead cleanup to remove shorter fragments, over 40% of inserts within each sequencing library were in the desirable size range (Fig. S1).

Contamination can be physically introduced in the laboratory or occur due to index misassignment errors during sequencing, resulting in a number of reads being incorrectly attributed to a sample. The presence of these contaminant reads can undermine several important inferences: estimations of viral load (in particular distinguishing low viral load from aviraemia), detection of drug resistant minor variants, and the inference of transmission direction using within-host phylogenetics. We identified and blacklisted contaminant reads using the previously described routine "phyloscanner clean" in the *phyloscanner* package (Fig. S2A). Out of the total set of HIV reads obtained from all samples, 1.2% of reads were blacklisted (median 6 reads per sample, mean 16 reads). As expected, the majority of contaminant reads were found in samples that had very few total HIV reads (Fig. S2B). To validate the blacklisting procedure, we looked at reads within the *pol* gene, which contains the majority of drug resistance mutations. In "spike-in" experiments, where known fractions of contaminant reads were introduced and then recovered, "phyloscanner clean" correctly blacklisted 262 out of 274 contaminant reads, giving an overall sensitivity of 95.6%. The distribution of the spiked-in reads over the 50 samples is shown in Fig. S2C. Of the 291,815 noncontaminant reads, 291,742 were correctly identified, giving an overall specificity of over 99.9%.

The cost of implementing a high-throughput virus genomics system will vary by setting. In our laboratory in Oxford, the reagent, consumables, and labor costs of the entire assay, from frozen blood to final data, is approximately 30 GBP in 2020, 3 times lower than the WHO budget recommendations for HIV pol sequencing in low-income settings (24). Costs were reduced by concentrating total nucleic acid extractions to allow library preparations with one-quarter reagent volumes without losses in sequencing sensitivity (Fig. S1). With a throughput of 10,000 to 15,000 samples per year, 30 GBP per sample covers the salary of a UK technician processing 350 samples per week. Laboratory set-up costs (ground rental, equipment, and maintenance costs) are not included in this calculation.

veSEQ-HIV yields quantitative viral loads. Viral load is the concentration of virus in a sample and is usually measured with highly standardized and regulated clinical assays using quantitative PCR to amplify both the material to be tested and spiked internal standards of known viral load. Viral load tests are essential for rapid detection of resistance-associated treatment failure, but are expensive and not a part of routine care in many low-income countries.

In a previous study of hepatitis C, we found that in contrast to amplicon-based sequencing, veSEQ was quantitative, in that total Illumina read-pairs correlated with clinical viral loads (25). To confirm that veSEQ was similarly quantitative for HIV, we performed both clinical viral load measurements and veSEQ-HIV sequencing on 146 specimens. Figure 2A shows the correlation between the routine clinically validated viral load and number of viral fragments recovered during sequencing, along with the R^2 value (0.89). This correlation was robust over a wide range of viral loads (Fig. 2B) that includes the quantifiable limit of the clinical assay ($<$50 copies/ml).

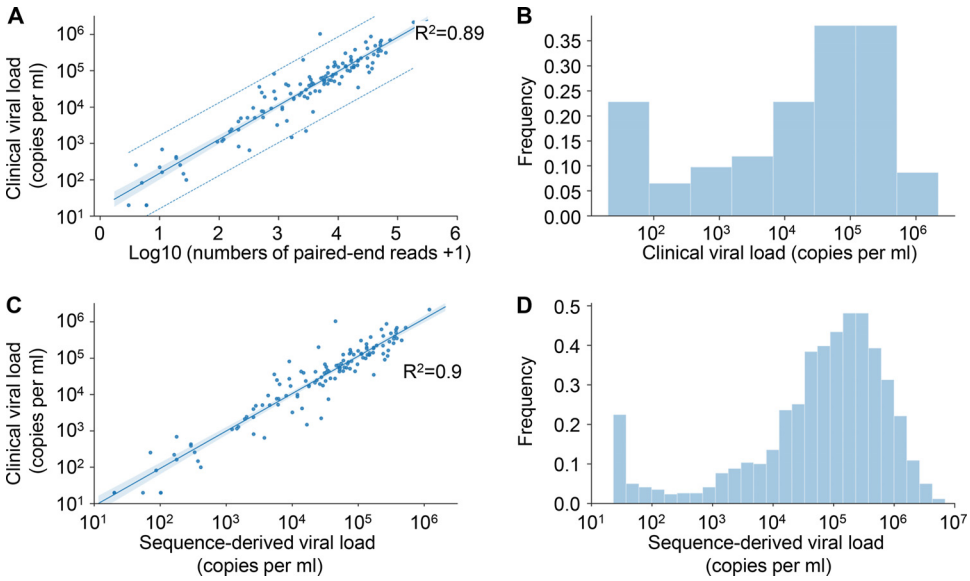

FIG 2 Viral load is calculated from the number of sequencing reads. (A) The data and linear regression model estimates for the viral load standards. The narrow shaded area is the 95% confidence interval for the regression curve, and the dashed lines are 95% prediction intervals for measurements. The mean squared error of prediction was 0.324 $\log_{10}$ copies/ml. (B) Distribution of independently measured clinical viral loads in a subset of 146 samples used to assess model performance. (C) Relationship between the clinical viral load and the sequence-derived viral load from the model shown in panel A for these 146 samples. (D) Frequency of sequence-derived viral load estimates for all 1,620 samples.

The relationship between number of reads and viral load was linear on a log-log scale with a slope of 0.83. This corresponds to some nonlinearity on a linear scale, consistent with some loss of information at high viral loads, possibly due to saturation effects or erroneous bioinformatic-compression of distinct reads into single "deduplicated" reads (which is expected, by chance, at very high sequencing depth). This does not affect the use of the number of viral fragments to infer viral loads, since the relationship is well described mathematically. We therefore defined "sequence-derived log viral load" as the linear transform of the log number of deduplicated sequence fragments (Fig. 2C). The lower limit of detection was approximately 50 copies/ml. We calculated the sequence-derived viral load for all sequenced samples using this transformation and characterized the population distribution (Fig. 2D). This distribution was bimodal, with the minor peak at very low viral load, corresponding to individuals with HIV read counts below the quantifiable limit of conventional assays. In line with procedures used to calibrate clinical viral load assays, a serial dilution of inactivated cultured virus was included in each run to ensure the quality of the assay, guard against batch effects, and quantify rates of contamination between samples.

veSEQ-HIV is unbiased with respect to viral genotype. Specificity for all known HIV subtypes circulating in Zambia was achieved using a probe-based, rather than a primer-based, amplification step (Fig. 3). HIV subtypes were inferred by sequence similarity to HIV reference genomes from the Los Alamos HIV database or by using the REGA HIV-1 subtyping tool (26). The predominant subtype was C, for which 86% (1,282/1,498) of samples yielded complete genomes. Eighteen nonsubtype C complete genomes included subtypes A (A1 and A2), D, G, and J, as well as the subtype B standards, demonstrating good probe affinity across HIV diversity. Given that partial genotypes are relatively harder to genotype correctly, it was unsurprising that 68% of ungenotyped sequences were incomplete (13/19), and those that were complete had features suggesting recombination.

Assay sensitivity and associations between viral load, read depth, and genome coverage. One of our aims was to ensure that veSEQ-HIV generated whole HIV genomes for the majority of samples within the range of viral loads observed in this

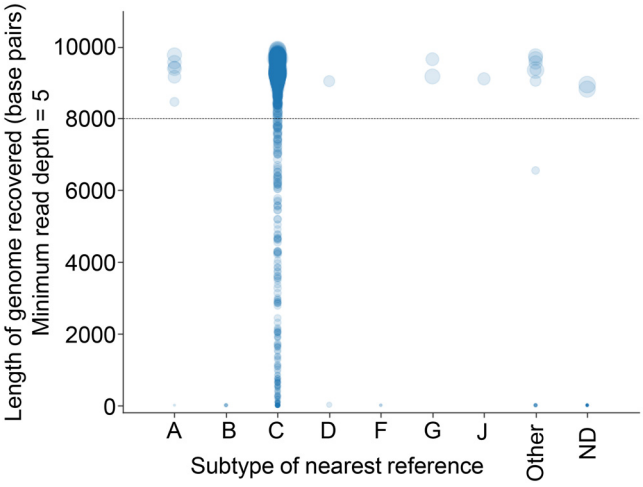

FIG 3 veSEQ-HIV is both sensitive and specific. Figure shows the length of recovered HIV genome for all sequenced samples. We consider a position in the genome to be accurately determined when the read depth is at least five. The category "Other" consists of potential intersubtype recombinants. Quantitative standards (HXB2, subtype B) are included in all sequencing runs, but are not displayed in this analysis.

population. The length of the recovered consensus sequence depends on the minimum read depth required to make a consensus call at each genomic position. We defined "read depth" as the number of mapped reads covering each position in the genome after removal of PCR duplicates. The point at which reads consistently matched the sample consensus saturated at a depth of five reads; we took this as our threshold for reliably inferring a consensus (Fig. S3). This may be a conservative estimate given that sequencing uninfected plasma (a negative control included in every run) resulted in no HIV reads after our multistage removal of contamination artifacts. However, we sought to produce not only accurate whole-genome consensus sequences, but also sufficient read depth for analyses of within-host diversity and characterization of low-frequency drug resistance mutations.

Whole HIV genomes, defined as having a sequence length over 8,000 bp with a minimum read depth of five deduplicated reads, were obtained from all 1,204 samples with a viral load greater than 10,000 copies/ml and from 1,297/1,424 samples (91%) with a viral load greater than 1,000 (Table 1). The lowest viral load for which a whole genome was obtained was 4,300 copies/ml and 97% of samples above this threshold produced a whole genome (Fig. 4A). The majority of commercially available HIV-genotyping tests require a viral load of over 1,000 copies/ml. In this data set, 6% of samples had viral loads within the range of 1,000 to 4,300 copies/ml; at this range, the average length of genome covered was 4,172 bp (Fig. 4A).

Higher viral loads in general resulted in higher read depth and therefore in greater coverage across the genome. Figure 4B shows the dependence of this success rate on sequence-derived viral load in more detail. Sigmoid functions (fit to the data with least-squares) indicate the viral load thresholds above which at least 8,000-bp genomes tend to be recovered: these are between 1,000 and 10,000 copies/ml, depending on the desired depth of reads supporting the consensus. Partial genomes were frequently obtained from samples with viral loads between 100 and 1,000 copies/ml (Fig. 4B).

TABLE 1 Numbers of samples processed using the sequencing pipeline and near-full genomes obtained (>8,000 bp), stratified by sequence-derived HIV-1 viral load (VL)

VL range (sequence derived)	Samples sequenced	Near-full-length genome
$<10^2$	126	0
10^2–10^3	68	0
10^3–10^4	220	93
$>10^4$	1,204	1,204

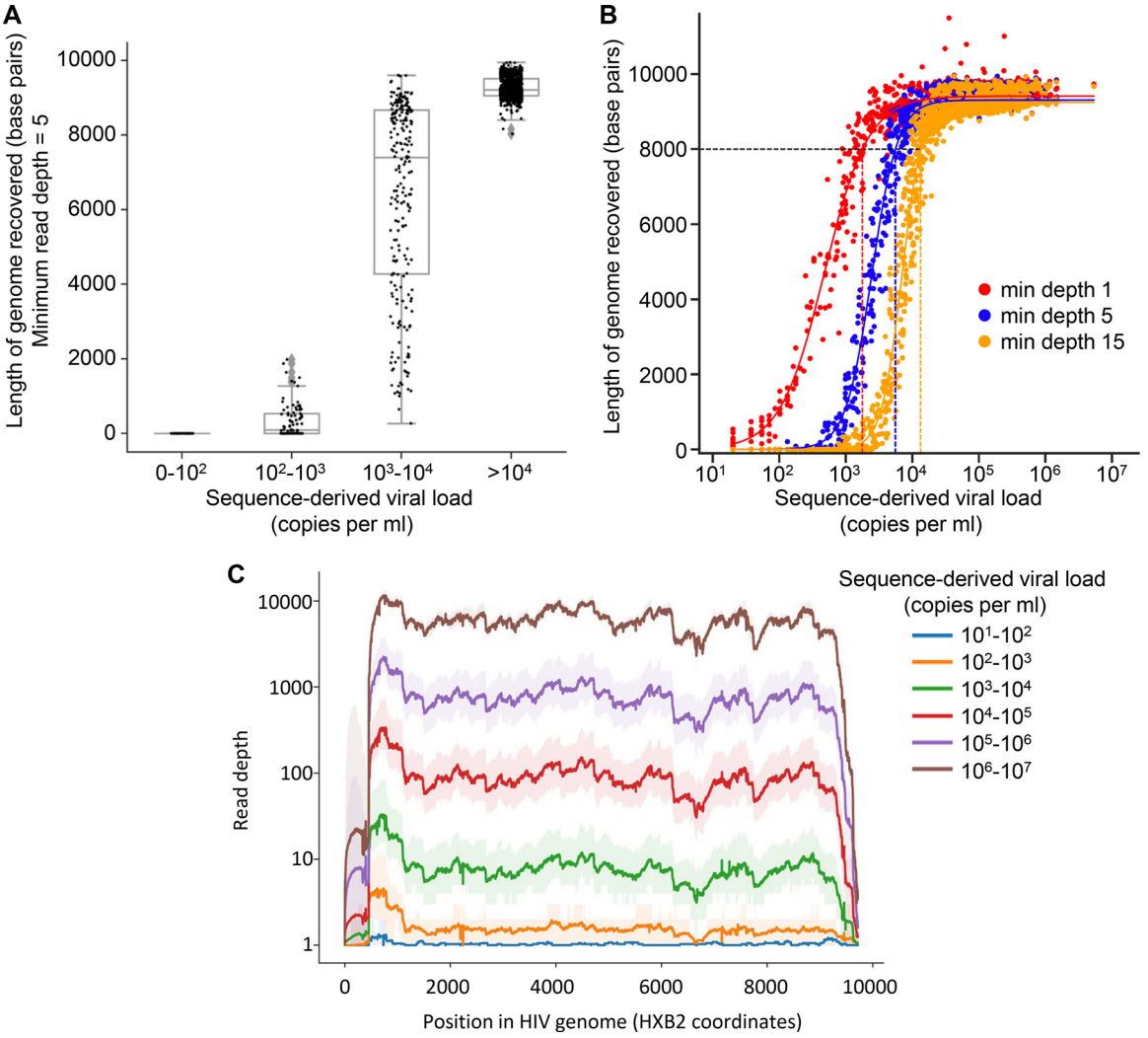

FIG 4 Sequencing success is influenced by viral load. (A) The length of the HIV genomes reconstructed by *shiver* software, from paired-end Illumina reads, stratified by log viral load, showed reproducible whole-genome coverage for samples with sequence inferred viral loads of >4 $\log_{10}$ copies/ml and near-complete coverage for the majority of samples with VL between 3 and 4 $\log_{10}$ copies/ml. (B) The viral loads at which genome coverage exceed 8 kb with minimum depth thresholds of 1 read, 5 reads, and 15 reads (after removal of PCR duplicates) are shown by the intercepts of curves fitted using a sigmoid function. (C) The median (thick lines) and 95th percentile range (ribbons) of read-depth observed across the genome are shown for all samples, grouped by sequence-derived viral load.

The patterns of read depth were reproducible between individuals, with similar patterns of high and low coverage across the genome (Fig. 4C). Importantly, we did not observe a drop-off in coverage below five reads to be systematically associated with particular parts of the genome (Fig. S4).

veSEQ-HIV provides drug resistance information on consensus and minority variant levels. The quantitative nature of the veSEQ-HIV pipeline and its ability to identify and remove contamination artifacts are useful properties for characterizing drug resistance mutations at low frequency. In accordance with previously published guidance on generating drug resistance inferences from next-generation sequence data (27), we implemented a simple algorithm, based on the *HIVdb* classification system (Stanford, US), to predict overall susceptibilities to antiretroviral drugs from the resistance mutations identified on individuals reads, after cleaning with *phyloscanner*. A novel output of this approach was a detailed description of all mutations and combinations of mutations linked to within-host phylogenetic information that *phyloscanner* uses to infer transmission. Figure 5 provides representative examples of two transmission pairs, for which the direction of transmission had been systematically determined

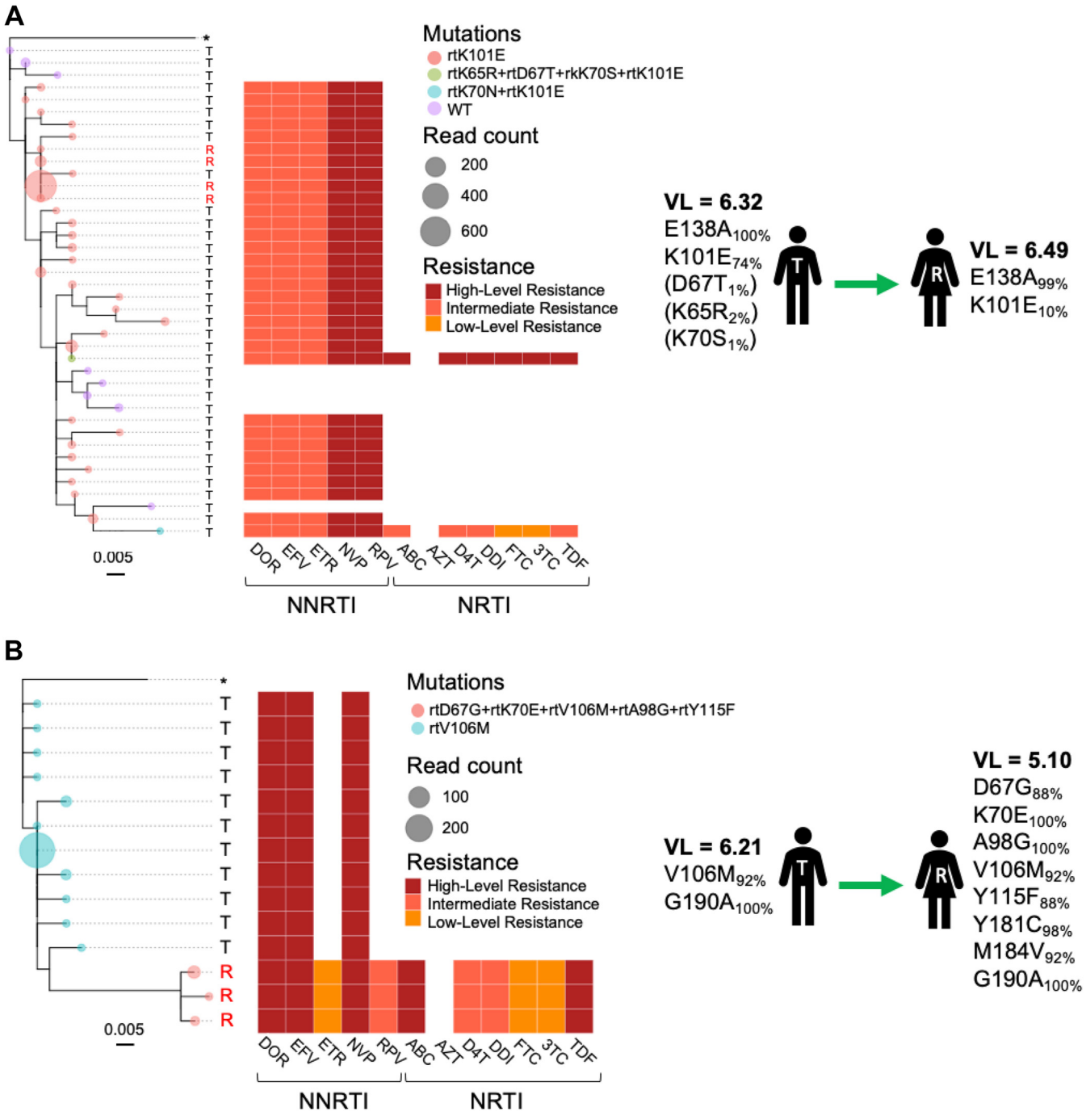

FIG 5 Within-host phylogenetic trees of Illumina reads spanning drug resistance sites in *pol*. *Phyloscanner* software performs ancestral state reconstructions of phylogenetic trees generated from Illumina reads in "windows" across the genome in order to identify pairs consisting of transmitters (T) and recipients (R). Phylogenetic trees of reads spanning drug resistance mutations sites in *pol* are shown for two inferred transmission pairs (A and B). Tree tips (circles) are colored by the combinations of drug resistance mutations observed for each unique taxon and scaled to total read counts within each taxa (after removal of PCR duplicates). Heatmaps report the predicted drug susceptibilities for each read using the Stanford HIVdb classification. Sequence-derived viral loads (log 10 RNA copies/ml) and the complete list of resistance mutations with associated frequencies, observed across entire genomes, are shown for each individual. Mutations observed at frequencies below 5% are shown in parentheses.

from ancestral-state reconstructions of multiple phylogenies of reads, performed in sliding windows across the genomes. Figure 5A depicts an example where a subclade of virus carrying the NNRTI resistance mutation K101E was transmitted to a female recipient. In the same transmission pair, subpopulations of wild-type susceptible virus and dual-class NRTI/NNRTI resistant virus (K65R/D67T/K70S/K101E and K70N/K101E)

were not transmitted to the recipient, probably because they were detected in the transmitter at low frequency (<5%). In another transmission pair (Fig. 5B), V106M and G190A mutations were detected in the female recipient along with a number of additional mutations (D67G, K70E, A98G, Y115F, Y181C, and M184V) that were not found in the male transmitter, suggesting these additional mutations were acquired after the inferred transmission event. Consistent with this finding, the female recipient reported prior knowledge of her HIV-positive status and previous use of ART, although she was not on treatment at the time of sampling. Both individuals were sampled within 2 months of each other, and in the same health care facility.

DISCUSSION

We have developed, optimized, and validated veSEQ-HIV, a fast, robust, cost-effective, and high-throughput laboratory and computational process for recovering complete HIV genomes, estimating viral load, detecting ART drug resistance mutations, and constructing transmission networks. The method has been shown to work with 1,620 genetically diverse samples collected from 10 Zambian clinics participating in HPTN 071-2 (PopART phylogenetics), producing whole genomes from >90% of samples with viral loads of >1,000 copies/ml. The assay works with residual plasma taken from routine CD4 cell count testing obtained in field laboratory conditions, without introducing undue contamination or degradation of the samples or the need for additional blood draws.

Our method has several advantages over previous high-throughput approaches (9). First, our probes were designed using an algorithm proven to tolerate levels of virus diversity even greater than that observed for HIV (e.g., HCV) (12), and are therefore expected to be unbiased with respect to the range of HIV variants commonly found in the region. Abbott Laboratories has recently reported on a similar method, developed in parallel to ours, which they show works across a wider panel of reference genomes (28, 29). Second, because our quantitative method minimizes the biases involved in PCR and computationally controls for contamination, our estimates of the frequency of minority genetic variants are likely to be more robust. Third, veSEQ is cost-effective. In our laboratory in Oxford, the reagent and consumables cost of the entire assay, from frozen blood to final data, is approximately 30 GBP in 2020, less than a fifth of the cost of the 2015 WHO budget for generating a full-genome sequence and viral load result (24). Our costing includes a technician salary, but not the initial costs of setting up a laboratory (equipment etc.)

The detection and quantification limits of veSEQ-HIV are comparable with those of clinical viral load assays (40 to 100 copies/ml), and inclusion of reference standards shows quantification is reproducible between runs. Sequencing from direct virus-PCR was previously shown to be less quantitative than veSEQ because of PCR saturation effects and because, unlike veSEQ, template resampling cannot be corrected with bioinformatic methods (for example, *PICARD MarkDuplicates*). Sequencing viral amplicons can be made quantitative with unique molecular identifiers (UMIs) that barcode single cDNA templates (30). In future work, we will evaluate whether addition of UMIs offers any additional benefit to the quantitativity and data quality of veSEQ-HIV. Here, we report an R^2 value of 0.89 in a comparison with the Roche AmpliPrep TaqMan system, which is well within range of reported R^2 values between commonly used clinical viral loads (0.80 to 0.94) (31). Additionally, *phyloscanner clean* provides a solution for "decontaminating" NGS data by removing low-frequency artifacts, such as index misassignments and PCR recombinants.

The throughput of veSEQ-HIV is suitable for large-scale public health applications. In our research setting, a single technician is able to process 360 samples per week. Routine combination testing to provide information on viral suppression, drug resistance, and transmission in near real time is feasible with veSEQ-HIV. This could prove useful as drug resistance surveillance is scaled up to guide and monitor new interventions, including preexposure prophylaxis, long-acting antiviral drugs, and alternative treatment regimens. High-resolution characterization of transmission events could

augment precision public health programs and focused responses to local outbreaks. However, we caution that patient groups should be regularly consulted on the ethical use of this technology, to provide maximum benefit while minimizing risks to individuals (32).

There remain important limitations to our approach. While we have validated our methods to minimize contamination and provide quality control tools to detect mix-ups as quickly as possible, the risk of large-scale mix-ups increases with higher throughput. This should be mitigated with sample barcoding and sample tracking. Second, veSEQ-HIV is not licensed for clinical viral load, genotyping, or drug resistance testing. However, as part of the HPTN-078 study, drug resistance mutations detected by veSEQ-HIV were concordant with those detected by the FDA-accredited HIV genotyping test, ViroSeq. This study also validated viral load estimates against the Abbot RealTime Assay and found that veSEQ-HIV obtained complete drug resistance information 93.3% of the time in samples with viral loads of $>$ 5000 RNA copies per ml (33).

The veSEQ-HIV protocol is tuned for high-throughput applications, and so is ideally suited for laboratories that process a large number of samples. Capital investments are modest, and the protocol is simple for technicians to adopt. However, maintenance and supply issues could be problematic in low- and middle-income countries, where the need is greatest. In such settings, centralized laboratory infrastructure could serve a number of districts. The computational component of the method is currently optimized for our local cluster infrastructure and will be streamlined and made platform-independent. Our current aim is for clinical accreditation of a complete laboratory and bioinformatics pipeline, operated by a single technician, data/lab manager, and clinical microbiologist, with remotely provided technical support, training, and quality assurance.

Future areas for improvement might include increasing automation, reducing initial capital expenditure costs, and reducing the reliance on regular supply chains of consumables. We did not explore the extent to which the bait capture step could be shortened or simplified; such improvements would further simplify the implementation of our method and would be needed to achieve a high-throughput protocol that could turn around sequence data in a single working day. Extending the length of individual sequences to capture whole viral haplotypes would improve applications in epidemiology and pathogenesis research.

The method can easily be adapted to study other RNA viruses, panels of RNA viruses, and even DNA and RNA viruses together, without loss of sensitivity (34) (preprint). Sequencing several pathogens simultaneously is achievable at minimal increased cost.

In summary, veSEQ-HIV is a cost-saving high-throughput protocol that, with current technologies, produces a sequence-derived viral load, a high-resolution drug resistance genotype, and data that can be used to provide highly granular insights into HIV epidemiology. The method has proven robust to field conditions in Zambia and carries no additional testing burden for patients. Sequencing will provide insights into the outcome of the HPTN 071 PopART trial (35) and, in our view, should be routinely performed in epidemiological and intervention studies of pathogenic viruses.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 3.9 MB.

ACKNOWLEDGMENTS

This work was sponsored by the National Institute of Allergy and Infectious Diseases (NIAID) under cooperative agreement numbers UM1-AI068619, UM1-AI068617, and UM1-AI068613 and funded by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR), the International Initiative for Impact Evaluation (with support from the Bill & Melinda Gates Foundation), NIAID, the National Institute of Mental Health (NIMH), and the National Institute on Drug Abuse (NIDA). E.B. was funded by the Medical Research

Council UK and the Oxford NIHR Biomedical Research Centre and is an NIHR Senior Investigator.

We thank Monique Andersson and the John Radcliffe Hospital, Oxford, Clinical Microbiology Department for assistance with viral load testing.

We acknowledge the support of the HPTN 071 (PopART) study team and Zambian Ministry for Health.

Sequencing was supported by the Oxford Viromics initiative (Paul Klenerman, David Bonsall, and Rory Bowden) and the Oxford Genomics Centre (With thanks to John Broxholme, Lorne Lonie, Angie Green, Jerome Nicod, and David Buck). Sample and data collection have been supported by the PANGEA HIV consortium, funded by the Bill & Melinda Gates Foundation.

REFERENCES

1. UNAIDS. 2017. 90-90-90—an ambitious treatment target to help end the AIDS epidemic. http://www.unaids.org/en/resources/documents/2017/90-90-90.
2. Akullian A, Bershteyn A, Jewell B, Camlin CS. 2017. The missing 27%. AIDS 31:2427–2429. https://doi.org/10.1097/QAD.0000000000001638.
3. Gupta RK, Jordan MR, Sultan BJ, Hill A, Davis DHJ, Gregson J, Sawyer AW, Hamers RL, Ndembi N, Pillay D, Bertagnolio S. 2012. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. Lancet 380:1250–1258. https://doi.org/10.1016/S0140-6736(12)61038-1.
4. Haile-Selassie H. 2017. WHO HIV drug resistance report. https://apps.who.int/iris/bitstream/handle/10665/255896/9789241512831-eng.pdf?sequence=1.
5. Wertheim JO, Pond SLK, Forgione LA, Mehta SR, Murrell B, Shah S, Smith DM, Scheffler K, Torian LV. 2017. Social and genetic networks of HIV-1 transmission in New York City. PLoS Pathog 13:e1006000. https://doi.org/10.1371/journal.ppat.1006000.
6. Weber J, Volkova I, Sahoo MK, Tzou PL, Shafer RW, Pinsky BA. 2019. Prospective evaluation of the Vela Diagnostics next-generation sequencing platform for HIV-1 genotypic resistance testing. J Mol Diagn 21:961–970. https://doi.org/10.1016/j.jmoldx.2019.06.003.
7. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, Gall A, Cornelissen M, Fraser C, Fraser, STOP-HCV Cosortium, The Maela Pneumococcal Collaboration, The BEEHIVE collaboration. 2018. PHYLOSCANNER: inferring transmission from within- and between-host pathogen genetic diversity. Mol Biol Evol 35:719–733. https://doi.org/10.1093/molbev/msx304.
8. Dennis AM, Herbeck JT, Brown AL, Kellam P, de Oliveira T, Pillay D, Fraser C, Cohen MS. 2014. Phylogenetic studies of transmission dynamics in generalized HIV epidemics: an essential tool where the burden is greatest? J Acquir Immune Defic Syndr 67:181–195. https://doi.org/10.1097/QAI.0000000000000271.
9. 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. https://doi.org/10.1128/JCM.01516-12.
10. Blanquart F, Wymant C, Cornelissen M, Gall A, Bakker M, Bezemer D, Hall M, Hillebregt M, Ong SH, Albert J, Bannert N, Fellay J, Fransen K, Gourlay AJ, Grabowski MK, Gunsenheimer-Bartmeyer B, Günthard HF, Kivelä P, Kouyos R, Laeyendecker O, Liitsola K, Meyer L, Porter K, Ristola M, van Sighem A, Vanham G, Berkhout B, Kellam P, Reiss P, Fraser C, collaboration BEEHIVE. 2017. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. PLoS Biol 15:e2001855. https://doi.org/10.1371/journal.pbio.2001855.
11. Ratmann O, Wymant C, Colijn C, Danaviah S, Essex M, Frost SDW, Gall A, Gaiseitsiwe S, Grabowski M, Gray R, Guindon S, von Haeseler A, Kaleebu P, Kendall M, Kozlov A, Manasa J, Minh BQ, Moyo S, Novitsky V, Nsubuga R, Pillay S, Quinn TC, Serwadda D, Ssemwanga D, Stamatakis A, Trifinopoulos J, Wawer M, Brown AL, de Oliveira T, Kellam P, Pillay D, Fraser C. 2017. HIV-1 full-genome phylogenetics of generalized epidemics in sub-Saharan Africa: impact of missing nucleotide characters in next-generation sequences. AIDS Res Hum Retroviruses 33:1083–1098. https://doi.org/10.1089/AID.2017.0061.
12. Bonsall D, Ansari MA, Ip C, Trebes A, Brown A, Klenerman P, Buck D, STOP-HCV Consortium, Piazza P, Barnes E, Bowden R. 2015. ve-SEQ: robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. F1000Res 4:1062. https://doi.org/10.12688/f1000research.7111.1.
13. Hayes R, Ayles H, Beyers N, Sabapathy K, Floyd S, Shanaube K, Bock P, Griffith S, Moore A, Watson-Jones D, Fraser C, Vermund SH, Fidler S, HPTN 071 (PopART) Study Team. 2014. HPTN 071 (PopART): rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment—a study protocol for a cluster randomised trial. Trials 15:57. https://doi.org/10.1186/1745-6215-15-57.
14. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol 13:104. https://doi.org/10.1186/1472-6750-13-104.
15. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46.
16. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
17. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
18. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
19. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
20. Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D, Croucher NJ, Hall M, Hillebregt M, Ong SH, Ratmann O, Albert J, Bannert N, Fellay J, Fransen K, Gourlay A, Grabowski MK, Gunsenheimer-Bartmeyer B, Günthard HF, Kivelä P, Kouyos R, Laeyendecker O, Liitsola K, Meyer L, Porter K, Ristola M, van Sighem A, Berkhout B, Cornelissen M, Kellam P, Reiss P, Fraser C, BEEHIVE Collaboration. 2018. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with *shiver*. Virus Evol 4:vey007. https://doi.org/10.1093/ve/vey007.
21. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519.
22. Shafer RW, Jung DR, Betts BJ. 2000. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. Nat Med 6:1290–1292. https://doi.org/10.1038/81407.
23. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques 30:892–897. https://doi.org/10.2144/01304pf02.
24. World Health Organization. 2015. HIV drug resistance surveillance guidance: 2015 update. https://apps.who.int/iris/bitstream/handle/10665/204471/9789241510097_eng.pdf?sequence=1.
25. Thomson E, Ip CLC, Badhan A, Christiansen MT, Adamson W, Ansari MA,

Bibby D, Breuer J, Brown A, Bowden R, Bryant J, Bonsall D, Filipe ADS, Hinds C, Hudson E, Klenerman P, Lythgow K, Mbisa JL, McLauchlan J, Myers R, Piazza P, Roy S, Trebes A, Sreenu VB, Witteveldt J, STOP-HCV Consortium, Barnes E, Simmonds P. 2016. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. J Clin Microbiol 54:2470–2484. https://doi.org/10.1128/JCM.00330-16.

26. Pineda-Peña A-C, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, Gómez-López A, Camacho RJ, 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. https://doi.org/10.1016/j.meegid.2013.04.032.
27. Ji H, Enns E, Brumme CJ, Parkin N, Howison M, Lee ER, Capina R, Marinier E, Avila-Rios S, Sandstrom P, Van Domselaar G, Harrigan R, Paredes R, Kantor R, Noguera-Julian M. 2018. Bioinformatic data processing pipelines in support of next-generation sequencing-based HIV drug resistance testing: the Winnipeg Consensus. J Int AIDS Soc 21:e25193. https://doi.org/10.1002/jia2.25193.
28. Bonsall D, Golubchik T, Kosloff B, Limbada M, de Cesare M, Schaap A, Hall M, Wymant C, Macintyre-Cockett G, Brown A, Ansari MA, Floyd S, Hayes R, Fidler S, Fraser C. 2018 HIV genotyping and phylogenetics in the HPTN 071 (PopART) study: validation of a high-throughput sequencing assay for viral load quantification, genotyping, resistance testing and high-resolution transmission networking, p 696. *In*: International AIDS Society, abstract THAA0101.
29. Yamaguchi J, Olivo A, Laeyendecker O, Forberg K, Ndembi N, Mbanya D, Kaptue L, Quinn T, Cloherty G, Rodgers M, Berg M. 2018. Universal target capture of HIV sequences from NGS libraries, p 47–48. *In* International AIDS Society, abstract TUPEA003.
30. Zhou S, Jones C, Mieczkowski P, Swanstrom R. 2015. Primer ID validates template sampling depth and greatly reduces the error rate of next-generation sequencing of HIV-1 genomic RNA populations. J Virol 89:8540–8555. https://doi.org/10.1128/JVI.00522-15.
31. Swenson LC, Cobb B, Geretti AM, Harrigan PR, Poljak M, Seguin-Devaux C, Verhofstede C, Wirden M, Amendola A, Boni J, Bourlet T, Huder JB, Karasi J-C, Lepej SZ, Lunar MM, Mukabayire O, Schuurman R, Tomazic J, Laethem KV, Vandekerckhove L, Wensing AMJ, International Viral Load Assay Collaboration. 2014. Comparative performances of HIV-1 RNA load assays at low viral load levels: results of an international collaboration. J Clin Microbiol 52:517–523. https://doi.org/10.1128/JCM.02461-13.
32. Coltart CEM, Hoppe A, Parker M, Dawson L, Amon JJ, Simwinga M, Geller G, Henderson G, Laeyendecker O, Tucker JD, Eba P, Novitsky V, Vandamme A-M, Seeley J, Dallabetta G, Harling G, Grabowski MK, Godfrey-Faussett P, Fraser C, Cohen MS, Pillay D, Ethics in HIV Phylogenetics Working Group. 2018. Ethical considerations in global HIV phylogenetic research. Lancet Hiv 5:e656–e666. https://doi.org/10.1016/S2352-3018(18)30134-6.
33. Fogel JM, Bonsall D, Cummings V, Bowden R, Golubchik T, de Cesare M, Wilson EA, Gamble T, del Rio C, Batey DS, Mayer KH, Farley JE, Hughes JP, Remien RH, Beyrer C, Fraser C, Eshleman SH. 8 August 2020. Performance of a high-throughput next-generation sequencing method for analysis of HIV drug resistance and viral load. J Antimicrob Chemother https://doi.org/10.1093/jac/dkaa352.
34. Goh C, Golubchik T, Ansari MA, de Cesare M, Trebes A, Elliott I, Bonsall D, Piazza P, Brown A, Slawinski H, Martin N, Defres S, Griffiths MJ, Bray JE, Maiden MC, Hutton P, Hinds CJ, Solomon T, Barnes E, Pollard AJ, Sadarangani M, Knight JC, Bowden R. 2019. Targeted metagenomic sequencing enhances the identification of pathogens associated with acute infection. bioRxiv https://doi.org/10.1101/716902.
35. Hayes RJ, Donnell S, Floyd S, Mandla N, Bwalya J, Sabapathy K, Yang B, Phiri M, Schaap A, Eshleman SH, Piwowar-Manning E, Kosloff B, James A, Skalland T, Wilson E, Emel L, Macleod D, Dunbar R, Simwinga M, Makola N, Bond V, Hoddinott G, Moore A, Griffith S, Deshmane Sista N, Vermund SH, El-Sadr W, Burns DN, Hargreaves JR, Hauck K, Fraser C, Shanaube K, Bock P, Beyers N, Ayles H, Fidler S, HPTN 071 (PopART) Study Team. 2019. Effect of universal testing and treatment on HIV incidence—HPTN 071 (PopART). N Engl J Med 381:207–218. https://doi.org/10.1056/NEJMoa1814556.