Published in final edited form as:

J Virol Methods. 2009 December ; 162(1-2): 280–283. doi:10.1016/j.jviromet.2009.08.002.

The Efficiency of Single Genome Amplification and Sequencing is Improved by Quantitation and Use of a Bioinformatics Tool

David M. Butler^{a,*}, Mary E. Pacold^b, Parris S. Jordan^a, Douglas D. Richman^{a,c}, and Davey M. Smith^{a,c}

David M. Butler: dmbutler@ucsd.edu; Mary E. Pacold: mpacold@ucsd.edu; Parris S. Jordan: psjordan@ucsd.edu; Douglas D. Richman: drichman@ucsd.edu; Davey M. Smith: davey@ucsd.edu

^aDepartment of Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0679, La Jolla, CA, USA 92039

^bBioinformatics Graduate Program, University of California, San Diego, 9500 Gilman Drive, Mail Code 0679, La Jolla, CA, USA 92039

^cVA San Diego Healthcare System, San Diego, CA, USA, 92161

Abstract

Typically, population-based sequencing of HIV does not detect minority variants present at levels below 20-30%. Single genome amplification (SGA) and sequencing improves detection, but it requires many PCRs to find the optimal terminal dilution to use. A novel method for guiding the selection of a terminal dilution was developed and compared to standard methods. A quantitative real-time PCR (qRT-PCR) protocol was developed. HIV RNA was extracted, reverse transcribed, and quantitated. A bioinformatics web-based application was created for calculating the optimal concentration of cDNA to use based on results of a trial PCR using the dilution suggested by the qRT-PCR results. This method was compared to the standard. Using the standard protocol, the mean number of PCRs giving an average of 30 (26-34, SD=3) SGA per sample was 245 (218-266, SD=20) after an average of 8 trial dilutions. Using this method, 135 PCRs (135-135, SD=0) produced 30 (27-30, SD=1) SGA using exactly two dilutions. This new method reduced turnaround time from 8 to 2 days.

Standard methods of SGA sequencing can be costly and both time- and labor-intensive. By choosing a terminal dilution concentration with the proposed method, the number of PCRs required is decreased and efficiency improved.

Keywords

Single Genome Amplification; quantitative real-time PCR; population-based sequencing; single genome sequencing; minority variant detection

© 2009 Elsevier B.V. All rights reserved.

*Corresponding author: (tel.) +1 858-552-8585 ext. 2673, (fax.) +1 858-552-7445.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Modern virology relies increasingly on the determination of nucleotide sequences of partial or complete viral genomes. Standard (or “bulk”) genetic sequencing utilizes population-based gene amplification by polymerase chain reaction (PCR) to create one sequence representing all the amplicons generated. When bulk sequencing reports a mixture of bases at a given position, this indicates diversity or the existence of minority variants within the genetic population, usually when minority variants comprise greater than 20% of the viral population (Gunthard et al., 1998). While this level of detection is adequate for a range of research applications, some virological studies require greater depth for characterizing minority strains (Johnson et al., 2008).

One method for increasing the detection of minority genetic variants within a sample population has been to use bacterial transformation for clonal amplification instead of population-based PCR. Clonal amplification can be both time- and labor-intensive, so newer ways to achieve deeper levels of detection have been developed. Single-Genome Amplification and Sequencing (SGA/S), for example, is a terminal dilution technique that has been used frequently in HIV research (Palmer et al., 2005; Shriner et al., 2004; Simmonds et al., 1990a; Simmonds et al., 1990b). This technique attempts to isolate a single molecule of viral DNA or copy DNA (cDNA) generated by reverse transcription (RT) of viral RNA through serial dilution testing. Specifically, DNA or cDNA is diluted serially over a range of concentrations, and the concentration at which $\leq 30\%$ of reactions contain amplifiable cDNA may be expected, assuming a Poisson distribution, to yield product generated from a single template in approximately 80% of those samples. By using a single molecule of DNA or cDNA as the template for amplification and sequencing, the risk of nucleotide misincorporations or template switching introduced during PCR amplification is reduced (Fang et al., 1998; Meyerhans et al., 1990; Palmer et al., 2005; Shriner et al., 2004; Simmonds et al., 1990a; Yang et al., 1996), and with repeated sampling of the viral population, SGA/S can obtain a greater than 20% level detection of diversity within a sampled viral population.

A major obstacle to the greater use of SGA/S is the time and cost involved in determining the appropriate dilution of DNA or cDNA to use. For example, SGA/S has been used to identify rare drug resistance mutations in HIV, where cDNA was diluted serially over a wide range of concentrations and then real-time PCR was used to identify the dilution at which only 3 of 10 samples had measurable product (Palmer et al., 2005). An alternative approach, involving persons infected recently with HIV (Salazar-Gonzalez et al., 2008), estimated the appropriate concentration of cDNA to use on the basis of blood viral loads. The concentration of cDNA was adjusted then by trial and error until a dilution yielded PCR product in $\leq 30\%$ of samples. Both SGA/S approaches improve the level of detection of minority variants within a viral population compared to bulk sequencing; however, both techniques require substantial amounts of viral template, which may be an important factor when the quantity of the clinical specimen is limited. To determine more efficiently the optimal dilution to use for SGA/S, use of quantitative real time PCR to measure directly the concentration of DNA or cDNA followed by use of a bioinformatics application to guide the dilution of cDNA to use for SGA/S is proposed.

2. Materials and Methods

2.1 Collection and Processing of Specimens

Blood plasma from persons infected with HIV in studies approved by the Human Research Protection Program of the University of California, San Diego, was used for these experiments. Blood was collected in acid-citrate-dextrose tubes by venipuncture and processed within 2

hours. Blood plasma was aliquoted, frozen, and stored at -80°C until processed for molecular studies.

2.2 Quantitation, Extraction, and Reverse Transcription of HIV-1 RNA

Blood viral loads were determined by extracting and quantifying HIV-1 RNA from 500 µL of blood plasma using the Amplicor HIV-1 Monitor kit (Roche Molecular Systems Inc., Alameda, CA, USA). Viral RNA was extracted from a separate aliquot of the same sample using the ViroSeq v.2.0 HIV genotyping system (Applied Biosystems, Foster City, CA, USA). For samples with viral loads above one million copies per milliliter, 250 µL of plasma were used; otherwise, 500 µL were used. Reverse transcription of the extracted HIV-1 RNA to cDNA was performed using RETROscript kits (Applied Biosystems, Foster City, CA, USA) using manufacturers' protocols. Specimens of cDNA were stored at minus 20° C until further use.

2.3 Calculation of Expected cDNA Dilution

2.3.1 Calculation of cDNA Concentration—The expected concentration of cDNA after reverse transcription is given by the formula:

$$\text{BPVL (copies} \times \text{mL}^{-1}) \times \text{PV (mL)} \times [\text{EV (mL)}]^{-1} \times \text{RV (mL)} \times [\text{FV (mL)}]^{-1},$$

where:

BPVL = blood plasma viral load for the specimen
 PV = plasma volume of sample used for reverse transcription
 EV = volume into which extracted RNA has been eluted
 RV = volume of RNA elution used in reverse transcription
 FV = final volume of reverse transcription reagents plus RV

2.3.2 Calculation of Target cDNA Dilution—The standard protocol for SGA/S (Birmingham, 2007) begins with an estimation of the amount of DNase-free water necessary to add to the sample of cDNA to dilute its concentration to approximately 10 copies per microliter (c/uL) and then dilutes this concentration by three-fold three times. This results in a range of dilutions for testing with hypothetical concentrations of 10 c/uL, 3.3 c/uL, 1.1 c/uL, and 0.4 c/uL. Each dilution is used in 16 wells for PCR, and if a dilution yields product in 4 wells, then that dilution is used for a full 96-well plate. If none of the dilutions yield 4 positive reactions, then more or less DNase-free water is added to alter the concentrations of cDNA, and the experiment is repeated until the right dilution is found based on 4 positive reactions.

2.4 Measurement of cDNA Concentration

Real-time quantitative PCR was performed by the Center for AIDS Research Genomics Core at the University of California, San Diego. HIV-1 RNA copies were quantified in a TaqMan-based approach as described previously (Heid et al., 1996). The forward primer (5'-TACAGTGCAGGGGAAAGAATA - 3'), which corresponds to nucleotides 4809-4829 of HXB2, the reverse primer (5' - CTGCCCCCTTCACCTTTCC - 3'), which corresponds to nucleotides 4957-4974 of HXB2, and the probe sequence (5' - TTTCCGGTATTACAGGGACAGCAG - 3'), corresponding to HXB2 nucleotides 4896-4922, were made (Integrated DNA Technologies Inc., Coralville, IA, USA) with specificity to the p31 integrase domain of *pol* (Rousseau et al., 2004). TaqMan standards were derived from a linearized, full-length HIV-1 clone, pNL-EX (courtesy of Dr. Yoshiharu Miura, Tohoku University Graduate School of Medicine, Sendai, Japan) in dilutions ranging from 1×10^6 copies/reaction to 20 copies/reaction. Each reaction consisted of 5 µL of HIV-1 standard

template or sample cDNA and 12.5 µL of 2X Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA). Primers and probe were present in final concentrations of 200 nM and 900 nM, respectively. All amplifications, including negative controls, were performed in duplicate with the ABI 7900HT Sequence Detection System (Applied Biosystems) using cycling parameters of 50° C for 2 min then 95° C for 10 min followed by 45 cycles of 95° C for 10s then 60° C for 1 min.

2.5 Determination of Actual cDNA Dilution

2.5.1 PCR of cDNA Dilutions

2.5.1.1 Dilution Procedure: The cDNA template used in these reactions was diluted to a concentration believed to approximate 0.4 copies/µL on the basis of either the method described in 2.3 (above) or by using the empirically obtained concentration of cDNA by qRT-PCR. Reactions were performed in parallel on a 96-well plate, and the second-round products were electrophoresed on 1% agarose gels to assess the fraction of positive reactions for a given dilution of cDNA. When greater than 30% of reactions were positive, cDNA was further diluted, and the amplification repeated until no more than 30% were positive. In cases where no reaction yielded product, a less dilute concentration of cDNA was used as template, and the amplification was repeated until no more than 30% of reactions were positive.

2.5.1.2 PCR Procedure: Nested polymerase chain reactions were performed using 10 µL of diluted cDNA template added to 40 µL of reaction mixture for the first round. The reaction mixture consisted in 5.0 µL of 10X PCR Buffer containing magnesium chloride and 1.0 µL of 10 nM dNTP Mix (GeneAmp, Applied Biosystems, Foster City, CA, USA), 0.25 µL of Taq DNA Polymerase (Roche Diagnostics, Indianapolis, IN, USA), 31.75 µL of molecular grade water, and 1 µL of each of two 20 M primers, V3-F_{out} (5'-CAAAGGTATCCTTTGAGCCAAT-3') and V3-B_{out} (5'-ATTACAGTAGAAAAATTCCCCT-3'). The 50 µL samples were heated to 95° C for two minutes and then subjected to 35 cycles of 30s at 95° C followed by 30s at 50° C followed by 60s at 72° C. After this, the samples were heated to 72° C for 10 min and then held at 4° C until used.

The second round PCR utilized 5µL of the first round product as template added to 45 µL of reaction mixture for a total volume of 50 µL. This reaction mixture consists of the same reagents, but the volume of molecular grade water is increased to 36.75 µL. For this round, the primers used are V3-F_{in} (5'-GAACAGGACCAGGATCCAATGTCAGCACAGTACAAT-3') and V3-B_{in} (5'-GCGTTAAAGCTTCTGGGTCCCCTCCTGAG-3'), but the thermal cycling parameters are the same as for the first round.

2.5.2 Bioinformatics Application for Interpreting Experimental Result—An application was developed that uses the Poisson distribution to calculate the real concentration of the terminal dilution (D) based on the number of positive PCR reactions (P) and the total number of PCR reactions run (T).

According to the Poisson distribution,

$$P(X=x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Where

$P(X=x)$ =probability of getting x template copies in one PCR reaction

λ =average number of cDNA template copies in each PCR reaction

X =number of cDNA template copies in one PCR reaction (random variable $[0, 1, 2, \dots]$);

Thus

$$P/T = P(X>0) = 1 - P(X=0) = 1 - (\lambda^0)(e^{-\lambda})/0! = 1 - e^{-\lambda}$$

Therefore:

$$\lambda = \ln(1 - P/T)$$

$$D = \lambda/10;$$

i.e., the actual concentration of the terminal dilution, as the PCR protocol calls for 10 μ L template per reaction.

This application, which is called “SGS Calculator”, also uses the real terminal dilution concentration (D) combined with two other inputs (the putative concentration of the terminal dilution and the putative concentration of the cDNA) to determine the real concentration of the cDNA. This is accomplished by solving for the unknown quantity after setting the following proportion:

$$\frac{\text{Actual cDNA concentration [unknown]}}{\text{putative cDNA concentration [known]}} = \frac{\text{Actual terminal dilution concentration [D, known]}}{\text{putative terminal dilution concentration [known]}}$$

The user must input the number of positive PCR reactions (P) and the total number of PCR reactions (T) from a trial run as well as the putative concentration (C) of the cDNA sample, as determined by quantitative real-time PCR, and the putative dilution (D) of the sample that was used in the trial run. The outputs of the SGS Calculator are the actual dilution of cDNA in the user's trial run and the actual concentration of sample.

For example, suppose the putative cDNA concentration (C) was 37 cp/ μ L based on the value obtained using quantitative real-time PCR. The user would make a dilution of 0.04 cp/ μ L (D, approximately the optimal 0.036 cp/ μ L) of cDNA for template in a trial run PCR plate. If the user obtained 50 positive PCR reactions (P) out of the 95 in a plate (T), rather than the optimal 28 positives (which would represent a positivity of 29.5%), then the following data would be entered into SGS Calculator:

Positive Wells (P)=50
Total Wells (T)=95
Putative Dilution (D)=0.04

After pressing “**Compute**”, SGS Calculator reports that the actual dilution of cDNA used was 0.075 cp/ μ L, and the actual cDNA concentration was 69.1 cp/ μ L. Both of these values are higher than their putative values, explaining why the number of positive PCRs was higher than

expected. A user would then use the calculated actual value for the cDNA concentration to make a new dilution of 0.04 or 0.036 cp/μL, which will give the desired percent PCR positivity.

This application is designed for large PCR sample sizes, i.e., the total number of PCR reactions run (T) should be at least 95. SGS Calculator is increasingly inaccurate for lower values of T. The application is implemented as a Java applet and currently available at <http://sgscalculator.ucsd.edu/>

3. Results

Using the standard SGA/S protocol, the mean number of PCRs required to obtain an average of 30 (range: 26-34, SD: 3) single genome amplification products per sample was 245 (range: 218-266, SD: 20) after an average of 8 trial dilutions. With the use of qRT-PCR and the bioinformatics tool, 135 PCRs (range: 135-135, SD: 0) produced 30 (range: 27-30, SD: 1) single genome amplification products per sample using exactly 2 dilutions. The turnaround time for generating SGA product for sequencing was reduced from 8 days using the standard approach to 2 days with the new method (Figure 1).

4. Discussion

To improve efficiency, a new method for determining the appropriate dilution of cDNA to use for SGA/S is proposed. Employing qRT-PCR to quantitate the nominal copies of cDNA after RT can lessen the observed discrepancies between the theoretically calculated and empirically determined optimal cDNA concentration to use for end-point dilution testing. The reasons for these discrepancies include: 1) a particular specimen may contain a concentration of HIV-1 RNA that is outside the dynamic range for which the viral load assay has its optimal accuracy; 2) the number of freeze/thaw cycles a specimen undergoes will affect the integrity of viral RNA available for participation in reverse transcription; 3) extraction of viral RNA from blood plasma may not capture all of the RNA measured by the viral load assay; 4) reverse transcription of RNA to cDNA is less than 100% efficient regardless of the procedure used, the number of freeze/thaw cycles, or the accuracy of the viral load assay. Furthermore, depending upon the type of research being performed, the amount of clinical material available for study can be a limiting factor with the standard method of SGA/S. If viral populations from compartments other than the blood (e.g., cerebrospinal fluid) are being characterized, then the quantity of sample limits the number of trials that may be performed searching for the optimal dilution to use for SGA/S. Using the methods proposed here, SGA/S was accomplished using on average only 2 rather than 8 trial dilutions. Although some of the original cDNA was used for the qRT-PCR test itself, less is used for this purpose than for a typical trial dilution in standard SGA/S. In conclusion, the method proposed here will increase the efficiency of the SGA/S procedure. This can reduce cost by decreasing the amount of reagents and labor involved, and it also may allow for application of this research tool to a broader range of investigations, as the amount of clinical material used for determining the optimal dilution is less than required previously.

Acknowledgments

We acknowledge the University of California San Diego, Center for AIDS Research Genomics Core Laboratory (Director, Dr. Christopher Woelk; Grant number, 5P30 AI36214) and the San Diego Veterans Medical Research Foundation. We thank Josue Santiago-Perez for assistance with the SGS Calculator website. This work was supported by grants AI69432, AI043638, MH62512, MH083552, AI077304, AI36214, AI047745, AI74621 from the National Institutes of Health and the California HIV/AIDS Research Program RN07-SD-702.

References

- Birmingham, MBaSCotUoAa. Standard Operating Procedure for Single Genome Amplification of HIV-1 Envelope. 2007.
- Fang G, Zhu G, Burger H, Keithly JS, Weiser B. Minimizing DNA recombination during long RT-PCR. *J Virol Methods* 1998;76:139–48. [PubMed: 9923748]
- Gunthard HF, Wong JK, Ignacio CC, Havlir DV, Richman DD. Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. *AIDS Res Hum Retroviruses* 1998;14:869–76. [PubMed: 9671215]
- Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res* 1996;6:986–994. [PubMed: 8908518]
- Johnson JA, Li JF, Wei X, Lipscomb J, Irlbeck D, Craig C, Smith A, Bennett DE, Monsour M, Sandstrom P, Lanier ER, Heneine W. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PLoS Med* 2008;5:e158. [PubMed: 18666824]
- Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. *Nucleic Acids Res* 1990;18:1687–91. [PubMed: 2186361]
- Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Dewar RL, Metcalf JA, Hammer S, Mellors JW, Coffin JM. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol* 2005;43:406–13. [PubMed: 15635002]
- Rousseau CM, Nduati RW, Richardson BA, John-Stewart GC, Mbori-Ngacha DA, Kreiss JK, Overbaugh J. Association of levels of HIV-1-infected breast milk cells and risk of mother-to-child transmission. *J Infect Dis* 2004;190:1880–8. [PubMed: 15499546]
- Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JE, Swanstrom R, Haynes BF, Athreya GS, Korber BT, Sharp PM, Shaw GM, Hahn BH. Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single Genome Amplification and Sequencing. *J Virol*. 2008
- Shriner D, Rodrigo AG, Nickle DC, Mullins JI. Pervasive genomic recombination of HIV-1 in vivo. *Genetics* 2004;167:1573–83. [PubMed: 15342499]
- Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990a; 64:5840–50. [PubMed: 2243378]
- Simmonds P, Balfe P, Peutherer JF, Ludlam CA, Bishop JO, Brown AJ. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J Virol* 1990b;64:864–72. [PubMed: 2296085]
- Yang YL, Wang G, Dorman K, Kaplan AH. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *AIDS Res Hum Retroviruses* 1996;12:303–6. [PubMed: 8906990]

Figure 1. Comparison of Standard and Proposed Methods

Using the standard method of single genome amplification (SGA), the mean number of PCRs to produce an average of 30 (26-34, SD=3) SGA per sample was 245 (218-266, SD=20) after an average of 8 trial dilutions. With the proposed method, 135 PCRs (135-135, SD=0) produced 30 (27-30, SD=1) SGA using exactly 2 dilutions.