The Sensitivity of HIV Deep Sequencing

Brendan B. Larsen, Wenjie Deng, Brandon Maust, Lennie Chen, Moon Kim, Dylan Westfall, Joshua Herbek, Eleanor Casey, Shyamala Iyer, Heather Bouzek, Hong Zhao, and James I. Mullins

University of Washington, Seattle, WA

Abstract

**Background**

Deep sequencing, primarily the 454/pyrosequencing platform, is increasingly being used to examine HIV viral populations, including for the detection of minor variants. We have examined how different variables impact the introduction of sequencing errors, thus affecting the sensitivity for detection of rare variants and accuracy of description of HIV quasispecies.

**Methods**

We performed pyrosequencing on multiple HIV male plasma control sequences, using a variety of PCR conditions and DNA polymerases to determine what factors affected error rates. We also sequenced a variety of plasmid mixtures at known rates to determine sensitivity and reproducibility of detection of minor variants.

**Results**

The polymerases used for PCR had the largest impact on the error rate and reproducibility of errors. PCR conditions, including the input number of templates and multiplexing, also had an effect.

**Discussion**

We found that the conditions of the PCR carried out prior to sequencing is the largest source of substitution errors following sequencing. Optimization of PCR conditions and enzymes can reduce background error rate and thus increase the depth of analysis of HIV quasispecies by deep sequencing methodologies.

**Background**

454/pyrosequencing has emerged as a powerful new technology to obtain vast amounts of data, allowing study of HIV population and population dynamics on a greater scale than was previously practiced. For example, pyrosequencing has been used to examine HIV duet (Campbell et al., 2011) and superinfection (Bui et al., 2011), drug resistance (Lalatade et al., 2012, Ganneva et al., 2011), viral tropism (Bunce et al., 2011), and viral escape from immune pressure (Fischer et al., 2010). These studies estimated variant frequencies and reported these estimates as representative of the true HIV population. However, as Swannemo and colleagues (2011) have noted, a possible bias in variant frequencies may occur due to PCR amplification. We examined this bias on a mixture of 16 HIV plasmids at different input levels that have undergone PCR using different polymerase combinations to determine the effect that PCR can contribute to the substitution errors and on the hypothesis for determining variant frequencies.

Although pyrosequencing is often called “deep sequencing,” it is unknown how deep into a heterogeneous population this technology actually probes. A 1% cutoff is often used to determine the relative abundance and variability of these variants as the most relevant determinants for the depth of this sequencing, the number of input templates that are sequenced should be considered. In the absence of independent sequencing experiments, and the number of templates sequenced, the background error rate, and therefore increases the reliable depth of deep sequencing.

**Methods**

We investigated the contribution to error rates of different DNA polymerases and PCR conditions. A known plasmid control was processed through our pyrosequencing amplification scheme (Fig. 1). All PCR conditions were evaluated on the same pyrosequencing plate to remove concerns about run-to-run variability. Briefly, 1st round single fragment or multiplex PCR were done on linearized plasmid-p4.2 DNA template. All enzyme combinations were run on the same pyrosequencing plate. Advantage 2 was the DNA polymerase used in several of the 2 round of PCR, the enzyme on the right is the enzyme that was used in the 2nd round of PCR.

**Results**

Figure 1 shows the difference between substitution error rates resulting from changes in PCR conditions. Effective error rates differ for all variants, but only those variants showed a statistically significant difference. Multiplying the first round PCR increased the error rate (p<0.001) compared to subtracting single fragments (which were then pooled together for the second rounds). We did not test the difference in error between single fragment 2nd PCR and multiplexed 2nd round PCR.

The other PCR condition that led to a difference in substitution error was the copies of template present at the start of each PCR. ~72 copies were put into 4 PCR reactions and then pooled afterwards for a total of 300 copies. ~28 copies were put into 20 reactions for the same total number of copies. Interestingly, pooling 20 reactions resulted in a lower error rate than when 4 reactions were pooled (p<0.05). For the enzyme comparisons, Kapa HiFi was used for both rounds of PCR showed a lower error rate than any other combination of enzymes (Table 2 and Fig. 5). The enzyme choice for both the first and second rounds had a significant effect on the substitution error.

The upper limit of background substitution error determines the cutoff for which variants can be accurately detected. Using our in-house control pipeline (see Postrel et al) we corrected enzyme error for the Kapa HiFi enzyme. In Figure 4, the correction of the Kapa HiFi plasmid control sequence leads to a maximum allele-specific error rate of 0.005 (95% CI), vs. 0.12 for Advantage 2. Besides errors occurring during PCR, another determinant for the sensitivity of deep sequencing was the number of input template molecules. In the field, these numbers are often, and poorly estimated from plasma viral loads measured on a different and shorter template using different conditions for RNA-extraction DNA production and PCR. We created a mixture of HIV plasmids that were represented in different frequencies. This plasmid mixture underwent PCR and 454 sequencing using 3 different input template numbers (Figures 5). As expected, lower frequency errors dropped out at the lower copy input.

**Discussion**

Pyrosequencing is often used to detect and quantify low-frequency HIV variants in a population. An arbitrary 1% cutoff is often used to distinguish background error from variant peaks. However, this represents basically a detection of variants that would otherwise be expected by the use of the term “deep sequencing,” and represents a <1% greater than the standard, and much lower error rate for sequencing technology using multiple templates.

In our experiments, we were able to determine the relative variability of detectable variants using pyrosequencing technology, we evaluated several parameters that we thought might influence error rate. The main causes of substitution error in 454 sequencing was found to be the type of DNA polymerase used, and how much DNA was produced during the PCR reaction. The latter was inferred by the fact that lower error rates were found if the linearized plasmid was used for both rounds of PCR. The former was confirmed by Lin et al. (2011) who investigated the correction of high frequency HIV variants when using a 454 sequencing protocol.

**References**


Table 1. Average error for each enzyme combination

<table>
<thead>
<tr>
<th>Condition</th>
<th>f of matched reads</th>
<th>f of positives</th>
<th>% of substitutions</th>
<th>% of deletions</th>
<th>% of insertions</th>
<th>Total of errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantage 2</td>
<td>36/42 (0.86)</td>
<td>27/36 (0.75)</td>
<td>0.23</td>
<td>0.11</td>
<td>0.00</td>
<td>34.70</td>
</tr>
<tr>
<td>Advantage 2/2</td>
<td>20/27 (0.74)</td>
<td>15/20 (0.75)</td>
<td>0.25</td>
<td>0.11</td>
<td>0.03</td>
<td>34.80</td>
</tr>
<tr>
<td>Kapa HiFi</td>
<td>24/20 (0.70)</td>
<td>15/24 (0.62)</td>
<td>0.16</td>
<td>0.34</td>
<td>0.04</td>
<td>18.60</td>
</tr>
<tr>
<td>Kapa HiFi/Adv.2</td>
<td>36/42 (0.86)</td>
<td>28/36 (0.77)</td>
<td>0.19</td>
<td>0.10</td>
<td>0.01</td>
<td>17.50</td>
</tr>
</tbody>
</table>

Table 2. Average error for each enzyme combination

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Advantage 2</th>
<th>Advantage 2/2</th>
<th>Kapa HiFi</th>
<th>Kapa HiFi/Adv.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phusion/Adv.2</td>
<td>36/42 (0.86)</td>
<td>27/36 (0.75)</td>
<td>24/20 (0.70)</td>
<td>36/42 (0.86)</td>
</tr>
<tr>
<td>Phusion/Adv.2/2</td>
<td>20/27 (0.74)</td>
<td>15/20 (0.75)</td>
<td>20/27 (0.74)</td>
<td>20/27 (0.74)</td>
</tr>
<tr>
<td>Kapa HiFi</td>
<td>24/20 (0.70)</td>
<td>15/24 (0.62)</td>
<td>24/20 (0.70)</td>
<td>24/20 (0.70)</td>
</tr>
<tr>
<td>Kapa HiFi/Adv.2</td>
<td>36/42 (0.86)</td>
<td>28/36 (0.77)</td>
<td>36/42 (0.86)</td>
<td>36/42 (0.86)</td>
</tr>
</tbody>
</table>