HIV Quasispecies and Resampling

Human immunodeficiency virus (HIV) behaves as an evolving quasispecies in infected individuals. When the populations of genetic variants that make up this quasispecies are characterized, it is common practice to obtain a number of nucleotide sequences by amplifying a region of the viral genome from a small amount of infected tissue [typically 1 microgram (μg) of DNA] by using the polymerase chain reaction (PCR), cloning the amplified products, and selecting a number of clones for sequencing (*1*). However, if the number of target molecules in the original sample is low, it is unlikely that all the sequences will have been derived from different input templates, and hence estimates of genetic diversity will often be below the true value (*2*). The controversy [reviewed by Barton F. Haynes *et al.* (Articles, 19 Jan., p. 324)] over whether slow or fast progressors to AIDS achieve higher levels of diversity may be rooted in this phenomenon.

Suppose N target molecules are subjected to PCR so that each is amplified to K molecules. Subsequently, the PCR products are cloned with equivalent efficiency, and r clones are randomly selected for DNA sequencing. K is very large in a typical PCR relative to r; hence, the probability distribution representing the frequency of selecting one particular clone follows a binomial function. The average number of distinct clones among r sampled clones is given by

$$N\left[1 - \left(1 - \frac{1}{N}\right)^r\right] \qquad (1)$$

The probability of resampling at least one original template is given

$$1 - \frac{\prod_{i=1}^{r}(N - i + 1)}{N^r} \qquad (2)$$

For a given number of clones selected, the higher the input number of target molecules, the higher the average number of unique clones chosen and the lower the probability of resampling. The average number of unique templates that the selected clones represent can also be estimated by formula 1 if the input copy number is known.

Analysis of HIV quasispecies from a commonly sampled source, such as uncultured peripheral blood cells (PBMCs) that typically contain 1 to 100 copies of proviral templates per microgram of DNA (*3*), can result in a high probability of resampling, especially for cases in which the copy number is low, for example, in slow or nonprogressors (*4*). For example, if 10 clones are to be sequenced, 25 input templates would give 88% probability of resampling at least one of the original templates. The dilemma of resampling is also evident when other techniques, such as heteroduplex mobility assays (HMAs) and heteroduplex tracking assays (HTAs) (*5*), are applied to HIV and other biological samples with low copy numbers of target molecules.

To avoid the resampling problem, each clone to be sequenced could be obtained from an independent PCR, which could be an expensive and laborious approach. Alternatively, proper estimation of quasispecies diversity will require quantitation of target templates performed by endpoint dilution (*3*) or other methods, and consideration of the issue of resampling.

Shan-Lu Liu
Allen G. Rodrigo
Raj Shankarappa
Gerald H. Learn
Department of Microbiology, University of Washington, Seattle, WA 98195–7740, USA

Li Hsu
Ori Davidov
Lue Ping Zhao
Quantitative Genetic Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA 98104, USA
James I. Mullins
Departments of Microbiology and Medicine, University of Washington

References

1. J. M. Coffin, *Science* **267**, 483 (1995).
2. S. Wain-Hobson, *Curr. Top. Microbiol. Immunol.* **176**, 181 (1992).
3. P. Simmonds *et al.*, *J. Virol.* **64**, 864 (1990).
4. Y. Cao, L. Qin, L. Zhang, J. Safrit, D. D. Ho, *N. Engl. J. Med.* **332**, 201 (1995).
5. E. L. Delwart *et al.*, *Science* **262**, 1257 (1993).

Response: Liu *et al.* point out the inherent difficulties in determining quasispecies diversity, particularly under circumstances where the sample size of target molecules is low. We agree that this might in fact contribute to the discrepancies in the literature regarding the levels of viral diversity in HIV-infected slow versus fast progressors.

As pointed out by Liu *et al.*, the sample size may play a critical role in evaluation of HIV quasispecies diversity. It is conceivable in reports of long-term nonprogressors that viral diversity was underestimated. Such technical constraints should be kept in mind when results are interpreted concerning patient populations with extremely low viral burdens.

Barton F. Haynes
Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA
Giuseppe Pantaleo
Anthony S. Fauci
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

■

Gene Therapy Chronology

Eliot Marshall states in his response to Robert J. Levine (Letters, 24 May, p. 1085) that I "told Levine in June 1995 that [I] did not foresee any possibility of a human trial," although this is clearly contradicted by Levine's statement "that the [*New York Times*] story was premature and exaggerated . . . but that as soon as [During] could foresee any possibility of extending [his] work to involve human subjects, [he] would contact [me] promptly."