HUMAN IMMUNODEFICIENCY VIRUSES

Too close for comfort

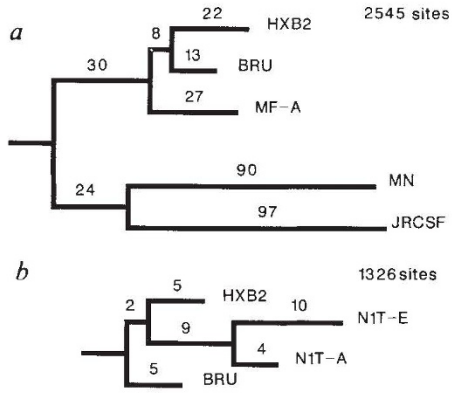
Simon Wain-Hobson and Gerald Myers

One of the hallmarks of the human immunodeficiency virus (HIV) is its genetic diversity. After 5 years of diligent DNA sequencing, the HIV database now contains genetic sequences from about 100 viral isolates, each of which displays a distinct genetic signature. With the exception of those obtained from a single source, or two closely related sources, pairs or groups of HIV isolates with genetic sequences differing by less than about 6 per cent are extremely rare. Against such a background it is hardly surprising that reports of HIV isolates whose sequences are strikingly similar but which supposedly originate from different patients should arouse special interest. One such case emerges from two papers appearing in the August issue of the *Journal of Virology* which provide evidence of two apparently independent, but essentially identical, HIV isolates. On the face of it these isolates would seem to call for a reevaluation of our understanding of HIV sequence heterogeneity. But do they?

First some basic facts. The genetic diversity of HIV is rooted in its replicative cycle. This involves three distinct steps: reverse transcription, DNA polymerization, and the synthesis of RNA from a DNA template (transcription). Any errors made by the polymerase enzyme during the first and third steps are not subject to proof reading, the result being pronounced sequence variability. A particular feature of the HIV genome is the existence of small duplications of 3–36 bases. These are most noticeable within the five hypervariable regions of the *env* sequence. In fact, the lability of these regions is such that genetic heterogeneity is not simply a feature of distinct HIV isolates but is also found among genomes within a single isolate. This suggests that if genetic comparisons are to be meaningful they should encompass a description of the plethora of viral genomes which, together, make up an HIV isolate.

Both E. I. Golub *et al.* (*J. Virol.* **64**, 3654–3660; 1990) and M. Stevenson *et al.* (*J. Virol.* **64**, 3792–3803; 1990) are interested in the functional heterogeneity of several different HIV-1 molecular clones derived from two viral isolates known as N1T and MF. As reported in the new papers, these isolates are from two different patients. The authors' aim is to identify small and subtle changes in the sequences of the clones which account for phenotypic differences such as infectivity and replicative efficiency.

The papers go some way towards that goal, but their particularity lies in what is not said. Thus the genetic restriction maps of the five molecular clones derived from the MF isolate are virtually indistinguishable from those of the N1T isolate. In turn the four N1T restriction maps are strikingly similar to those of two other viruses, HTLV-IIIb and LAV-1, famous for being among the first HIV-1 isolates to be characterized in the mid-1980s. The clustering of MF-A, N1T-E, N1T-A, BRU and HXB2 is evident in a phylogenetic tree analysis of the DNA sequences reported in the papers (see figure).



Tree analysis produced by Swofford's PAUP program and based on *(a) env* sequences, and *(b) vif* plus LTR (long terminal repeat) sequences. Each analysis included the same set of 9 other North American sequences. Only the nearest neighbours MN and JRCSF are shown. BRU is from LAV-1 whereas HXB2 is from HTLV-IIIB. N1T-A/-E and MF-A are equidistant from both BRU and HXB2.

The similarity of these sequences runs counter to expectations. Indeed so much so, that when one realizes that both the N1T and MF isolates were derived in the same facility, N1T first and then MF, and that authors of both papers cosigned an earlier paper on the biological properties of the N1T isolate, the notion that MF could in fact be N1T, rather than an independent isolate, begins to take hold. By the same token, could the similarities between N1T and the prototypical HIV-1s — HTLV-IIIB and LAV-1 — reflect an extension of the chain? Can we be absolutely sure that N1T did not originate from LAV-1?

Although none of this detracts from the validity of the papers — both groups worked with molecular clones so there was no need for them to describe the inherent sequence variability of their isolates — it does deserve comment. But another paper should also be noted. Earlier this year, T. McNearney *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **87**, 1917–1921; 1990) reported sequence data derived from seven apparently independent HIV isolates from a virus donor and two recipients. What is astonishing is that these authors discovered only four variable nucleotides within 13 entire *env* sequences. And of course all sequences were co-linear throughout their hypervariable regions. There is nothing like this in the entire HIV sequence database, which now contains more than 1,000 sequences. A number of other features, notably that all the *env* products are functionally defective, make these data very hard to assimilate — they simply buck the trend.

So, either we have some awkward observations that threaten to overturn the established view of HIV genetic diversity, or else these sequences are not quite what they appear to be. At this stage, the possibility of viral contamination should not be ruled out. Contamination of cells is well known to all cell biologists. Thus, SIV contamination has already been described, as has contamination of genomic DNA by phages and plasmids. Added to this is the threat posed by contaminant amplification through the ubiquitous use of the polymerase chain reaction (PCR). One of us (S.W.-H.) has encountered more than a dozen cases of PCR contamination due to laboratory plasmids and M13 phages.

No one can say for sure whether the sequences of the clones derived from MF are in fact those of clones inadvertantly derived from an N1T contaminant. But this possibility remains, given that materials were exchanged between the two groups of authors.

Are such issues important? The purity and authenticity of biological reagents and data are always crucial, but especially so when the identity of an HIV-1 sequence is used as an indication of where the virus came from. The widely assumed genetic diversity of HIV isolates is currently being turned into the basis of a method for establishing unusual or rare transmission routes, a development that raises the stakes considerably. A recent investigation into the possible transmission of HIV from a dentist to a patient during an invasive oral procedure, for example, revolved entirely on the similarity of two HIV-1 sequences (*Morbidity and Mortality Weekly Report* **39**, 489–493; 1990).

We believe that as much attention and rigour must be used in the molecular characterization of HIV-1 isolates as is necessary in human DNA fingerprinting. To establish the identity of an HIV isolate, we suggest that a population of sequences should be sought. If a viral population is not observed, or if any two are uncannily similar, then the onus should be on the sequencers to establish the origin of the viral isolate and lay to rest any fears about contamination. Given the current state-of-the-art technology, those of us in AIDS research should not settle for less. □

Simon Wain-Hobson is at the Institut Pasteur, 28, Rue Du Dr Roux, 75724 Paris Cedex 15, France. Gerald Myers is at the Los Alamos National Laboratory, New Mexico 87545, USA.