DEPARTMENT OF HEALTH & HUMAN SERVICES

# Memorandum

20 September 1988

From

Dr. Gerald Myers
LHNCBC-NLM, Rm. BS-816
496-2478
seqanalysis

Subject

To

Dr. Robert Gallo
NCI, Bldg. 37, Rm. 6A09

We have not met but I assume you are familiar with the HIV Sequence Database and Analysis Project, of which Flossie and Steve are editors. Thanks to Mary Reitz and Vefia Franchini, we have been able to send out without delay several sequences.

From our earliest tree analyses, it was patently evident that the LAV and IIIB viruses had to have had a recent common ancestor. We have not suppressed this fact in our publications (or talks); neither have we commented upon it, much as we have been pressed to do so. The recent papers of Li et al. and of Yokoyama et al. in Mol.Biol.Evol. 5,313 and 5,237 respectively confirm the same general result although different computer algorithms were employed. All of these analyses have been limited by the total amount of sequence information analysed; thus while they have consistently clustered the French and NCI viruses together, they could not consistently articulate the branching order of the group so long as a limited number of sites over different gene sequences were being studied.

Because of the current interest in "swarming" (Temin's term), it has been advantageous to develop fine-structure analyses of this sort...high resolution trees, if you like. The work of Kahn and Shaw; of Goodenow and Wain-Hobson; and now of the Cloning and Sequencing Project funded by the AIDS Program all make it essential to closely track variants that have recently diverged from one another. We have, accordingly, initiated such computer analysis using the WMJ sequences from the earlier paper, the SIVMACs, some unpublished data from the above-mentioned Pasteur study, and the IIIB/BRU sequences. The latter become of further interest due to their prominence; e.g. their continuing value as evolving sequences (the infected labworker, the recent discrepancy between the Reitz HXB2 and the Ratner HXB2, etc.).

By including all of the available gene sequences in a single analysis for the IIIBs, it is actually possible to define the branching order of the variants to a high degree of statistical precision. There is no doubt but that it shows the LAV source of the IIIB viruses: the NL43 clone of the BRU isolate is the oldest sequence; the published BRU follows it; the IIIBs follow thereafter (see enclosed figure).

We are, of course, checking and rechecking these results very carefully; they hold across the several analyses we have run thus far. It seems only

right to apprise you of them given your stake in the outcome. I'd be glad to discuss the work with you, but I would urge you to independently consult with an expert or two in this field of sequence analysis.

If you have any sequence data or suggestions pertaining to this inquiry that may shed different light upon the picture, I would be most happy to try to integrate that. Sequences are like fingerprints; up to now, the "smudges" have precluded any definitive judgment about the origin of the so-called IIIBs, other than that they were very close relatives of the LAV (the closest other HIV1 sequenced to date is MN, as you can see from the enclosed tree). After this analysis, it would be extraordinarily difficult to explain the fingerprints any other way but how I have summarized. Should our thinking change in light of some new information, I will be sure to inform you.

Please let me know your thoughts. I don't expect you to act now on this in any particular way. Eventually the result will have to be put forward to the community as part of the inquiry into fine-structure variation: perhaps the matter requiring immediate attention along these lines is the apparent variation of the HXB2 through reclening (Reitz sequence vs. Ratner sequence).