Introduction

This Compendium is an annual printed summary of the data contained in the HIV sequence database. In these compendia we try to present a judicious selection of the data in such a way that it is of maximum utility to HIV researchers. Traditionally, we present the sequence data themselves in the form of alignments: a comprehensive alignment of a selection of all full-length genomes the database contains (a lot of LAI-like sequences, for example, have been omitted because they are so similar that they bias the alignment) of HIV-1/SIVcpz (Section 1) and a combined HIV-1/HIV-2/SIV whole genome alignment (Section 2); amino acid alignments for HIV-1/SIC-cpz, HIV-2/SIV, and SIVagm. The HIV-2/SIV and SIVagm amino acid alignments are separate because the genetic distances between these groups are so great that presenting them in one alignment would make them very elongated because of the large number of gaps that have to be inserted. As always, tables with extensive background information gathered from the literature accompany the whole genome alignments.

New this year are two-page-wide amino acid alignments for all HIV-1 proteins. The collection of whole-gene sequences in the database is now large enough that we have abundant representation of most subtypes (excluding H and J). For most other subtypes, and especially for subtype B, a large number of sequences that span entire genes were not included in the printed alignments to conserve space. A more complete version of all alignments is available on our website, `http://hiv-web.lanl.gov` Importantly, all these alignments have been edited to include only one sequence per person, based on phylogenetic trees that were created for all of them, as well as the literature. At the request of many users, we have re-inserted the consensus sequences for each subtype, unless there are fewer than 5 sequences representing it. In the alignments we have also included the 'Circulating Recombinant Forms', so far four mosaic genomes that have epidemiological significance (see the nomenclature chapter for more on CRFs). To date four mosaic genomes have been described. The annotation of the Nef alignment has been expanded based on the Nef review that can be found in the Reviews (Section 6). Finally, for all amino acid alignments we have decided to combine the annotation tables into one, because of the increasing redundancy in the separate tables. In addition to sequence information (accession numbers, references) the new table lists which regions of the sequence are represented in the alignments.

We have made an effort to bring the HIV-2/SIV and SIVagm alignments up-to-date as well. We have created an entirely new HIV-1/HIV-2/SIV alignment that is much improved over the previous version; it can be accessed via our website. Because of the frequency of redundant information, we have decided to merge the gene tables into one large table for each alignment section; we hope you will find these tables easy to use.

In the Reviews section, along with the previously mentioned chapter on HIV nomenclature, you will find a very clearly written and concise overview of the functions of Nef; a review of monkey and SIV phylogeny including an 'idealized monkey tree'; and an extensive discussion of all aspects of HHV8 (KSHV), the recently discovered herpesvirus associated with Kaposi's Sarcoma. In addition, we present updated versions of the customary reviews of coreceptor usage, drug resistance, and SIV/SHIV vaccine reagents. Reprints of all reviews are available from our website in the form of both HTML and PDF files.

As always, we are open to complaints and suggestions for improvement. With the effort that goes into producing these volumes, we sincerely hope they will be widely used by the research community. Inquiries and comments regarding the Compendium should be addressed to:

Dr. Carla Kuiken
Theoretical Division, T-10, MS K710, LANL, Los Alamos, NM 87545
Ph: (505)-665–6463; fax: (505)-665–3493; e-mail: `kuiken@t10.lanl.gov`