



**Figure 1. The complete HSV-1 genome includes two unique regions and two sets of large inverted repeats.**

(A) The full structure of the HSV-1 genome includes a unique long region (UL) and a unique short region (US), each of which is flanked by inverted copies of a large repeat, known as the terminal and internal repeat of the long region (TRL and IRL) and the short region (TRS and IRS). The gene content of each region (UL, US, TRL/IRL and TRS/IRS) is distinct, as shown in Fig. 3. Lengths of each region are marked below; these are drawn approximately to scale. A short cleavage and packaging sequence called a is located as a direct repeat at both genome termini (in TRL and TRS) and as an inverted repeat (a') where IRS and IRL overlap. (B), Since sequences originating from one copy of an inverted repeat could not be distinguished from sequences originating from the other copy, the data were assembled into a trimmed form lacking the terminal repeats TRL and TRS. GenBank accessions contain both a full-length and a trimmed version for each genome (see Methods for details).