# HIV-1 Nomenclature Proposal
## A Reference Guide to HIV-1 Classification

**D.L. Robertson[1], J.P. Anderson[2], J.A. Bradac[3], J.K. Carr[4], B. Foley[5], R.K. Funkhouser[6], F. Gao[7], B.H. Hahn[7], M.L. Kalish[11], C. Kuiken[5], G.H. Learn[2], T. Leitner[8], F. McCutchan[4], S. Osmanov[9], M. Peeters[10], D. Pieniazek[11], M. Salminen[12], P. M. Sharp[13], S. Wolinsky[14], and B. Korber[5,6]**

[1] *Dept. of Zoology, Univ. of Oxford, Oxford, UK*

[2] *Dept. of Microbiology, Univ. of Washington, Seattle, WA*

[3] *Division of AIDS, NIH, Bethesda, MD*

[4] *Henry M. Jackson Foundation, Rockville, MD*

[5] *Group T-10, Los Alamos National Lab., Los Alamos, NM*

[6] *Santa Fe Institute, Santa Fe, NM*

[7] *Depts. of Medicine and Microbiology, Univ. of Alabama at Birmingham, Birmingham, AL*

[8] *Dept. of Clinical Virology, Swedish Inst. for Infectious Disease Control, Solna, Sweden*

[9] *Dept. of Policy, Strategy and Research, UNAIDS, Geneva, Switzerland*

[10] *Laboratoire Retrovirus, IRD, Montpellier, France*

[11] *HIV/AIDS and Retrovirology Branch, CDC. Atlanta, GA*

[12] *Dept. of Infectious Disease Epidemiology, National Public Health Inst., Helsinki, Finland*

[13] *Institute of Genetics, Univ. of Nottingham, Nottingham, UK*

[14] *Dept. of Medicine, Northwestern Univ. Medical School, Chicago, IL*

## I. OVERVIEW

Globally circulating strains of human immunodeficiency virus type one (HIV-1) exhibit an extraordinary degree of genetic diversity, which may influence aspects of their biology such as infectivity, transmissibility and immunogenicity. Sequences derived from these HIV-1 strains have historically been classified on the basis of their phylogenetic relationships into groups and subtypes. However, the increasing complexities of newly derived HIV-1 sequences have indicated a need to re-evaluate the current nomenclature system. In September 1999, a meeting was held at the Santa Fe Institute, New Mexico, to discuss the shortcomings of the HIV-1 nomenclature now in use. The goal was to resolve ambiguities, while at the same time retain as much of the current nomenclature as possible, so as to avoid inconsistencies with the existing literature. A summary of the deliberations and resulting recommendations is described below.

## II. INTRODUCTION

A first attempt to classify HIV-1 sequences was to subdivide them into European/North American and African strains, as sequences derived from European and North American isolates formed a distinct cluster in phylogenetic trees, while strains from Africa separated into different lineages (Li *et al.* 1988; Myers *et al.* 1988). However, when additional sequences from other geographic regions became available, it was clear that this classification system was too limited. Instead phylogenetic analysis of envelope sequences revealed the existence of multiple phylogenetic clusters, or clades, which were approximately equidistant from one another. These clades were named subtypes A to F, with the prototypic "North-American/European" strains relabelled subtype B (Myers *et al.* 1992). Subsequently, five of these six env-based subtypes/clades (A, B, C, D and F, but not subtype E) were identified in phylogenies inferred from the gag region (Louwagie *et al.* 1993). In the following years four additional

subtypes, G to J, were characterized based on phylogenetic comparisons of partial sequences (Janssens *et al.* 1994; Kostrikis *et al.* 1995; Leitner *et al.* 1995; Louwagie *et al.* 1995). More recently, subtype F was reported to be comprised of sub-subtypes F1, F2 and F3 based upon gag and env phylogenetic comparisons (Triques *et al.* 1999), but after subsequent analysis of complete genomes, sub-subtype F3 was renamed subtype K (Triques *et al.* 2000). [The latter subtype "K"should not be confused with partial sequences recently designated as "k" which cluster as a sister clade to subtype D (Roques *et al.* 1999).] Collectively all of the HIV-1 subtypes group together to form a clade which has been termed group M for "main", to distinguish them from the HIV-1 group O (outlier) clade (Gurtler *et al.* 1994), and the recently discovered HIV-1 group N (non-M/non-O) clade (Simon *et al.* 1998).

The great majority of HIV-1 strains cluster consistently in phylogenetic trees, that is they fall into the same subtype (or group) regardless of which regions of their genomes are analyzed. However, it was recognized early on that a fraction of HIV-1 strains exhibit discordant branching orders in phylogenies inferred from different parts of their genomes (Robertson *et al.* 1995). This finding, along with the fact that all of these viruses originated from geographic regions where the same divergent sequence subtypes co-circulated, strongly suggested that these viruses were the product of recombination events. This propensity for HIV to recombine was not unexpected, given results from previous retrovirus research (Coffin 1979; Hu and Temin 1990a; Hu and Temin 1990b) and specific HIV studies (Li *et al.* 1988; Sabino *et al.* 1994; Diaz *et al.* 1995), and it is now well established that recombination is a relatively common occurrence among different strains of HIV (reviewed in Robertson and Gao 1998; Quinones-Mateu and Arts 1999). This is most obvious among members of different subtypes, but is also likely to occur among members of the same subtype, although current methods fail to reliably identify such intra-subtype recombinants. One of the more interesting, and epidemiologically important examples of recombinant strains, are the so-called "subtype E" viruses which are most prevalent in Thailand and neighbouring countries in Southeast Asia. Evidence that "subtype E" viruses might represent recombinants was inferred initially from phylogenetic studies of gag and env (McCutchan *et al.* 1992; Louwagie *et al.* 1993), and later from analyses of complete "subtype E" genomes (Carr *et al.* 1996; Gao *et al.* 1996). In the extracellular portion of gp120 and gp41, "subtype E" viruses have long been known to cluster as a distinct group, a finding which led to their initial classification as an independent subtype (Myers *et al.* 1992). In contrast, in regions such as gag and pol, all "subtype E" viruses fall within the subtype A radiation. Thus, "subtype E" appears to comprise a recombinant lineage between subtypes A and E, although a "clean" (non-recombinant) subtype E lineage has not been found leading to some debate about the recombinant status of subtype E (see Discussion).

Recent advances in long PCR technologies have made it possible to generate full-length HIV genomic sequences on a routine basis (Salminen *et al.* 1995a; 1995b; 1996). This again has influenced HIV-1 nomenclature, as it is now clear that recombination breakpoints frequently occur throughout the HIV genome. It also has become apparent that subtype assignments in the accessory gene region can be difficult. For example, all known subtype G strains are relatively more closely related to subtype A in the vif/vpr region. Whether this indicates an anomaly or a recombinant ancestry of subtype G remains unresolved (Carr *et al.* 1998b; Gao *et al.* 1998b). Recent studies have also shown that new subtypes cannot be assigned on the basis of subgenomic sequences only. For example, an HIV-1 isolate previously classified as subtype I on the basis of C2V3 sequences (Kostrikis *et al.* 1995) was found to be a complex recombinant comprised of subtypes A, G and regions that do not fall into any of the currently defined subtypes (Gao *et al.* 1998a; Nasioulas *et al.* 1999). Moreover, different non-contiguous regions of unknown subtype origin, initially all termed "subtype I", were later found to be closely related to either subtype H or K, or unclassified sequences representing at least one other unknown subtype (Salminen 1999). This case exemplifies the need for full-length genomic sequences for new subtype designations. Finally, some HIV-1 inter-subtype recombinants are spreading epidemically (McCutchan *et al.* 1999; Montavon *et al.* 1999; Nasioulas *et al.* 1999). These have been termed "circulating recombinant forms" (CRFs) (Carr *et al.* 1998a) to indicate that they represent strains that are contributing to the global epidemic.

The results of the deliberations at the Nomenclature Workshop and final recommendations are listed below.

## III. THE REVISED HIV NOMENCLATURE SYSTEM

### A. Four types of categories should be used to refer to the major HIV-1 lineages: groups, subtypes, sub-subtypes and CRFs

• **Groups** will continue to refer to the very distinctive HIV-1 lineages M, N and O (see Figure 1 and Gao *et al.* 1999 for a description of the relationship of these different HIV-1 lineages to SIVcpz strains). Group M includes the viruses that dominate the global epidemic and the sub-division of this group is the focus of this proposal. The groups were originally named M for main, O for outlier, and N for Non-M-Non-O (DeLeys90, Charneau94, Simon98).

• **Subtypes** will continue to refer to the major clades within group M. Examples of the subtype structure in phylogenetic trees, and the genetic distances within and between subtype sequences can be seen in Figures 1, 2 and 3.

• **Sub-subtype** will be used to refer to a distinctive lineage that is very closely related to a particular subtype lineage, and is not genetically distant enough to justify calling a new subtype. The previously recognized subtype, for example A, would thus be renamed sub-subtype A1 and the hypothetical newly identified lineage named sub-subtype A2. Examples of the sub-subtype organization in phylogenetic trees, and the genetic distances within and between subtype and sub-subtype sequences can be seen in Figures 1, 2, and 3.

• **Circulating Recombinant Form (CRF)** describes a recombinant lineage, that plays an important role in the HIV-1 pandemic. The CRF members must share an identical mosaic structure, *i.e.*, they are descended from the same recombination event(s). The mosaic genome structures and subtype compositions of the four currently recognized CRFs are schematically shown in Figure 4.

### B. General requirements for defining a new subtype, sub-subtype or CRF

In order to define a new subtype, sub-subtype or CRF, representative strains must be identified in at least three individuals with no direct epidemiological linkage. Three near full-length genomic sequences are preferred, but two complete genomes in conjunction with partial sequences of a third strain are sufficient to designate a new subtype, sub-subtype or CRF (to define a CRF, the partial sequence(s) must also confirm the CRFs mosaic structure).

### C. Guidelines for designating subtypes and sub-subtypes

A combination of phylogenetic and distance analyses should be used to define a new subtype or sub-subtype. A new subtype should be roughly equidistant from all previously characterized subtypes in all regions of the genome with a distinct pre-subtype branch similar to those of other subtypes (see Figures 1 and 2). Given the extent of sampling of strains from the HIV-1 pandemic it is unlikely that many more non-recombinant subtype-like lineages will be found. However, it is possible that as sampling continues, particularly from Central Africa, more lineages will be found that do not fit neatly into the current subtype system. Whether such lineages will ultimately confound the current classification system is difficult to predict at the present time.

A sub-subtype should be identified when a group of viruses is significantly more closely related to one particular subtype, *i.e.*, the new lineage should form a sister clade to the subtype in question. Such sub-subtypes were identified in the past but were misclassified. For example, it appears that in retrospect subtypes B and D should have been classified as sub-subtypes B1 and B2 or D1 and D2, rather than as separate subtypes (Figure 2), but for the sake of consistency with published literature subtypes B and D will not be renamed. Lineages that exhibit different relationships to other subtypes in different regions of their genome are potentially CRFs (see section F), and should be analyzed appropriately.
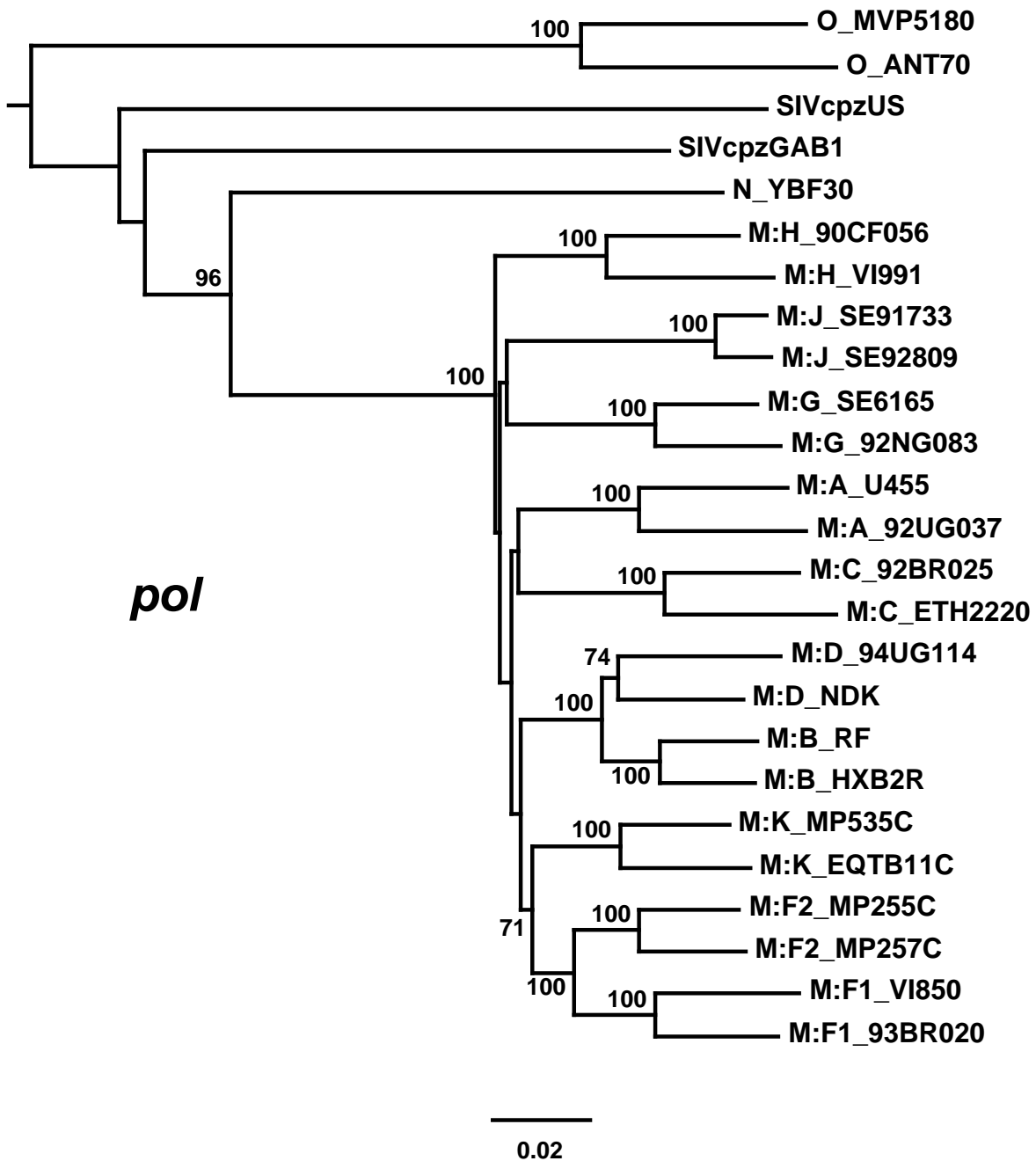
**Figure 1.** Phylogenetic relationships between representative strains of HIV-1 group M subtypes A-D, F1, F2, G, H, J and K, groups N and O, and SIVcpz (P.t.t.) inferred from pol nucleotide sequence comparisons. The pol region was taken from an alignment of representative near full-length strains. The alignment was gap-stripped and this mid-point rooted phylogenetic tree generated by neighbor-joining using DNADIST and NEIGHBOR from the PHYLIP package (Felsenstein 1993), with the F84 model, a transition:transversion ratio of 1.5 and empirical base frequencies. The numbers indicate percentage bootstrap replicates (of 1000) calculated using SEQBOOT, DNADIST, NEIGHBOR and CONSENSE from the PHYLIP package; values below 70% are not shown. The scale bar indicates 2% nucleotide sequence divergence.
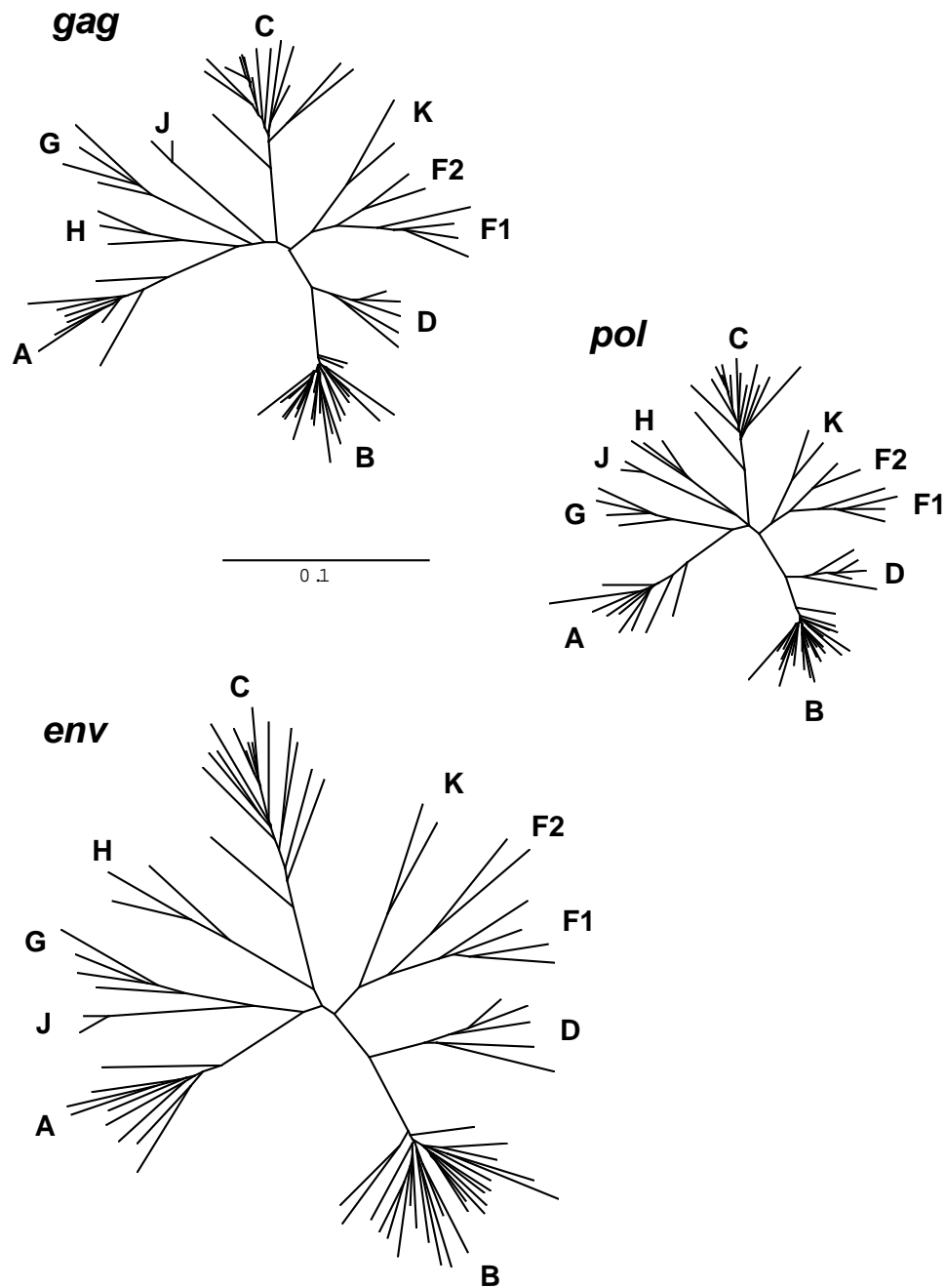
Reviews

**Figure 2**. Phylogenetic relationships between representative strains of group M subtypes A-D, F1, F2, G, H , J and K inferred from gag, pol and env nucleotide sequence comparisons. The gag, pol and env regions were taken from an alignment of representative near full-length strains, including only one sequence per infected individual. The alignments were gap-stripped and these unrooted phylogenetic trees generated by maximum likelihood using fastDNAml (Olsen et al. 1994), with the F84 model, a transition:transversion ratio of 1.5 and empirical base frequencies. The same strains are present in each of the trees. All three trees are drawn to the same scale, to facilitate comparison of the rates of evolution between genes. The scale bar indicates 10% nucleotide sequence divergence.

To aid researchers in distinguishing between subtypes and sub-subtypes, genetic distance analyses in conjunction with phylogenetic approaches may be useful. To develop these genetic distance guidelines, within- and between-class distances to ancestral nodes were compared in gag, pol and env trees (Figure 2), or else simple pairwise distance comparisons were compared, again, in gag, pol and env regions, for all available full-length sequences that could be unambiguously assigned to a non-problematic class. Figure 3 illustrates these within-subtype, between-subtype and between-sub-subtype distances. These distances can be used to judge whether a new lineage is more likely to be a new subtype or sub-subtype. For example, the plots show that sub-subtypes F1 and F2 have distances that are similar to subtypes B and D. Using different substitution models for the distance calculations did not affect the relationship between the distributions of sub-subtype and subtype distances. However, it has been shown that more realistic models are preferred for detailed evolutionary studies (Leitner *et al.* 1997), and the relationships between the distributions may change due to increased distances in the future. Thus, it should be emphasized that all distance analyses should be carried out using appropriate reference strains and backed up with detailed phylogenetic analyses.

To provide investigators with an easy screening method, a new tool based on genetic distance comparisons, designated the Subtype Distance Tool (SUDI), will be available at the Los Alamos National Laboratory (LANL) HIV Sequence Database website (http://hiv-web.lanl.gov). This online tool will enable a user to automatically create plots of pairwise comparisons similar to those in Figure 3 for strains of unknown subtype or sub-subtype classification. For example, the strains MP535C and EQTB11C that define subtype K (Triques *et al.* 2000), when entered into SUDI with reference strains, give the results that sub-subtype F2 is the closest in gag and pol while F1 is the closest in env. From this, the program calculates the distances from K to the closest lineage, and to the other subtypes, to see if these distances are in the between-subtype, between-sub-subtype or intra-subtype ranges. The results (not shown) indicate that subtype K is more distant from sub-subtypes F1 and F2 than these are from each other, but closer to F1 and F2 that to other subtypes, *i.e.*, in the range of subtype B and D distances. This relationship of subtype K to subtype F is confirmed by phylogenetic analysis (see Figure 2), indicating that subtype K should have perhaps been retained as sub-subtype F3. However, in Triques and co-workers analysis of complete genomes (Triques *et al.* 2000) subtype K did not consistently form a sister clade to subtype F, which is indicative of a possible recombinant ancestry of the K/F3 lineage. Since subtype K is now an established subtype in the literature its designation will not be changed.

### D. Naming of groups, subtypes and sub-subtypes

• The names of subtypes A-D, F-H, J and K will be retained; these appear to represent non-recombinant lineages relative to the known lineages (Salminen *et al.* 1996; Gao *et al.* 1998b; Laukkanen *et al.* 1999; Triques *et al.* 2000) with the possible exception of the ambiguous region in the accessory genes of subtype G (see section E). The subtype E designation will also be retained to refer to the putative non-subtype A regions in the A/E recombinants (see section E).

• Any future subtypes will continue to be named alphabetically, *i.e.*, L, M, N etc.

• Sub-subtypes will be named with a number following the subtype letter.

• The group M, N and O nomenclature will be retained; future groups will be named alphabetically, *i.e.*, P, Q, R etc. To distinguish between group M and subtype M (when/if the latter is characterized) the subtypes would be referred to as M:A, M:B.. M:M etc., when it is necessary to avoid confusion. Similarly, if future research identifies subtypes within groups N and O, these should similarly be labelled N:A, N:B etc. and O:A, O:B etc. Note, future subtype and CRF designations within groups N and O will have to meet the same criteria as have been established for group M.

• Until lineages have met the criteria required for a designation as a subtype, *e.g.*, when only partial sequences or less than three genomic sequences have been obtained (see section B), they will be labelled "U" for unclassified.

### E. The "problematic" subtypes E, G and I:

-Subtype E will now be referred to as CRF01_AE (see section F) to reflect its distinctive
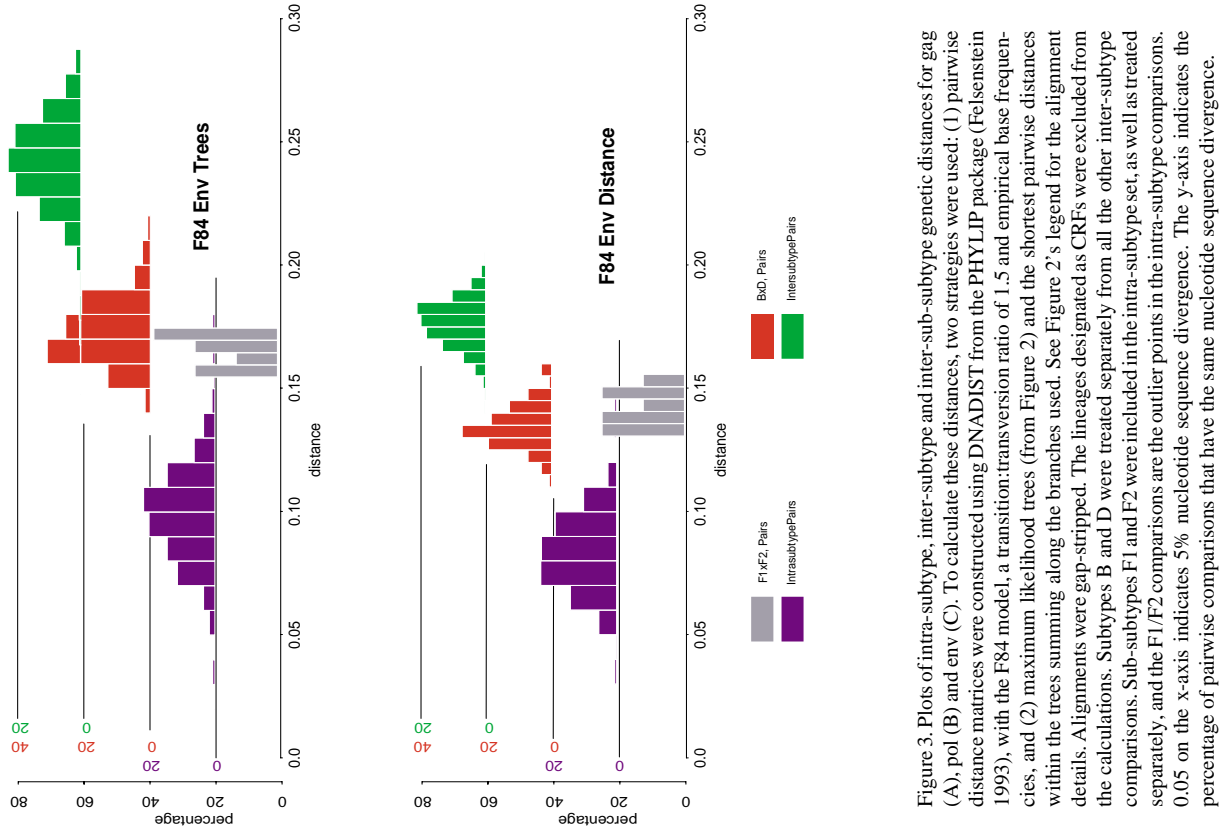
Figure 3. Plots of intra-subtype, inter-subtype and inter-sub-subtype genetic distances for gag (A), pol (B) and env (C). To calculate these distances, two strategies were used: (1) pairwise distance matrices were constructed using DNADIST from the PHYLIP package (Felsenstein 1993), with the F84 model, a transition:transversion ratio of 1.5 and empirical base frequencies, and (2) maximum likelihood trees (from Figure 2) and the shortest pairwise distances within the trees summing along the branches used. See Figure 2's legend for the alignment details. Alignments were gap-stripped. The lineages designated as CRFs were excluded from the calculations. Sub-subtypes F1 and F2 were included in the intra-subtype set, as well as treated separately, and the F1/F2 comparisons are the outlier points in the intra-subtype comparisons. 0.05 on the x-axis indicates 5% nucleotide sequence divergence. The y-axis indicates the percentage of pairwise comparisons that have the same nucleotide sequence divergence.

phylogenetic clustering in env relative to the rest of the genome, and the view held by the majority of the working group's participants that it is an A/E recombinant (see Discussion). Although the "E-like" segments in this CRF should strictly be called "U" according to the recommendations of this proposal, the E designation for the envelope region will be retained for historical consistency with the existing literature.

-As mentioned in the Introduction, all presently characterized subtype G strains share the same ambiguous subtype A relationship in their accessory region indicating a possible recombinant ancestry of this lineage (Carr *et al.* 1998a; Gao *et al.* 1998b). However, because of the lack of a "clean" subtype G parental representative strain, coupled with the short length of the region in question, the naming of the subtype G lineage will be retained rather than it be redefined as recombinant.

-The subtype I designation has been dropped from the nomenclature because the isolates from Cyprus and Greece (94CY032, PVMY and PVCH [the latter two strains are also named GR11 and GR84]), which were earlier classified as recombinants of A, G and a putative new subtype, I (Nasioulas *et al.* 1999) have, upon re-analysis with previously unavailable complete genome sequences, been revealed to be mosaics with regions associated with subtypes A, G, K , H and unclassified regions (Figure 4; Salminen 1999). Thus, subtype I will be removed from the genetic classification system of HIV strains, and the "I" regions will be relabelled as unclassified (U). The letter "I" will not be reused in the subtype nomenclature to avoid confusion with the existing literature.
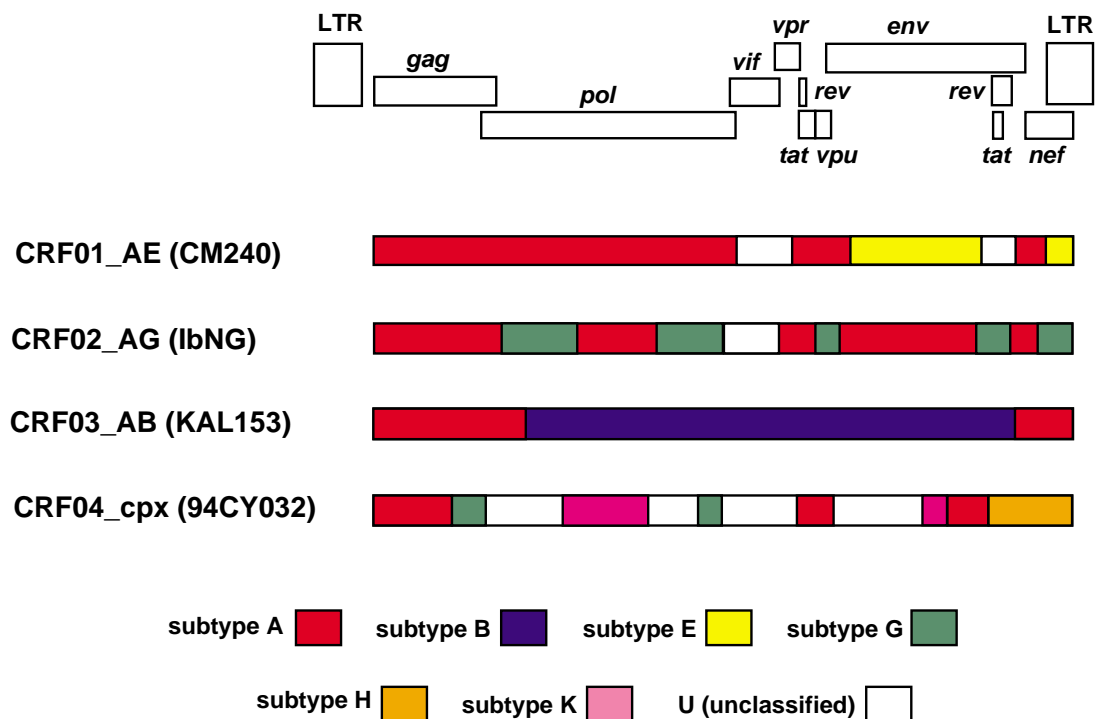


Figure 4. Mosaic genome structures of the four currently recognized circulating recombinant forms: CRF01_AE, CRF02_AG, CRF03_AB, and CRF04_cpx. An alignment of representative near full-length strains was used. This alignment was gap-stripped prior to all analysis. At the top of the figure, an HIV-1 genome map shows the position of each open reading frame in the gap-stripped multiple alignment. Below the genome map, each bar represents the mosaic pattern of a CRF. The different colours correspond to different subtype assignments. The white regions correspond to unassigned regions. The LTRs were not analyzed. The recombinant regions were inferred from diversity plots, bootscans (Salminen et al. 1995) and when possible from informative sites analysis (Robertson et al. 1995). The complete analysis can be found at http://grinch.zoo.ox.ac.uk/HIV/Figure_4.html. Also, see the primary publications mentioned in section F for each of these CRF's definitive recombination analysis.

**F. The CRF nomenclature**

Each circulating recombinant form will be given an identifying number, with letters (listed alphabetically) indicating the subtypes involved, *e.g*., CRF02_AG designates the IbNg-like strains circulating in Africa composed of genomic segments from subtypes A and G (Carr *et al.* 1998b). If more than three subtypes are found to make up the recombinant form, it would be designated as "cpx" (complex) rather than a list of the subtypes involved being given. The first complete genome sequenced from a CRF should be used as the reference strain. The four CRFs presently identified (Figure 4) are, in the order of their discoveries:

• CRF01_AE (reference strain CM240) which represents a putative subtype A/E recombinant form of HIV-1 which is spreading epidemically in Asia, but that originated from Central Africa (Murphy *et al.* 1993; Carr *et al.* 1996; Gao *et al.* 1996). In the future, putative recombinants with only one full-length "parental" subtype representative would be designated as being comprised of this subtype and unclassified regions (U). Under the new system CRF01_AE would be referred to as CRF01_AU because the putative "parental" non-recombinant E strain has not been found. But, as the "E" designation for the env region of these strains is very commonly used, renaming it would lead to confusion. Thus, the "E" designation will be retained.

• CRF02_AG (reference strain IbNg [Howard and Rasheed 1996]) which represents a subtype A/G recombinant form that is circulating in West and Central Africa (Carr *et al.* 1998b; Carr *et al.* 1999).

• CRF03_AB (reference strain KAL153) which represents a subtype A/B recombinant form that is circulating in Kaliningrad, primarily in injecting drug users (Liitsola *et al.* 1998; Salminen 1999). Circulation of this strain appears to have been accelerated by intravenous injection of a locally produced opiate contaminated with HIV infected blood.

• CRF04_cpx (reference strain 94CY032) which represents a Cypriot/Greek recombinant form that was previously classified as an A/G/I recombinant (Gao *et al.* 1998a; Nasioulas *et al.* 1999). This recombinant has recently been found to be an even more complex mosaic comprised of subtypes A, G, H, K and unclassified regions (Salminen 1999). Note that the "I" designation has been dropped from the nomenclature (see section E).

**G. Classification of genomic regions that do not cluster with the known subtypes**

Full-length genomes or partial regions of strains that form distinct lineages relative to the known subtypes, and do not meet the criteria for designating a new subtype or sub-subtype, will be labelled as unclassified (U). Segments within inter-subtype recombinant genomes, including segments within CRFs, for which the "parental" strain cannot be determined, will be labelled "U", *i.e*., a subtype can only be defined for a lineage inferred to be non-recombinant. A table including all of the strains designated "U" will be added to the subtype reference section at the LANL HIV Sequence Database website (http://hiv-web.lanl.gov) on the subtype reference alignments page.

**Notes:**

1.  Sequences that do not span the entire genome cannot be designated a new subtype, or CRF, but can still, of course, be assigned to an existing subtype, or CRF. However, such a designation refers only to the specific fragment sequenced, as the fragment may be embedded within a recombinant genome.

2.  Isolates representative of new subtypes, sub-subtypes and CRFs will be made available through the NIH AIDS Research and Reference Reagent Program (contact Opendra Sharma, osharma@niaid.nih.gov).

3.  We suggest that a simplified version of the WHO style nomenclature (Korber *et al.* 1994) be used to name all newly derived HIV strains, and include the following information: year and country of identification, unique laboratory identification, and clone number. For example, "99USunique_ID05" corresponds to a strain sampled in 1999 from a person in the US, unique_ID to the in-house laboratory designation, and 05 is the clone number. A list of two letter country codes can be found at the LANL HIV Sequence Database website (http://hiv-web.lanl.gov) on the DBSearch page. The two character code

"00" will be used for strains identified in the year 2000. Occasionally a person's residence and their probable place of infection are different; in these cases, the residence should be included in the strain name, and the place of infection recorded in the Features or Comments section of the GenBank/EMBL entry to aid epidemiological tracking. Indeed, it would be generally helpful if as much information as possible about a viral strain and its host could be included in the GenBank entry to aid comparative studies.

4. Authors who plan to describe new sub-subtypes, subtypes, or CRFs should consult with database staff prior to publication to avoid redundant namings and general inconsistencies with this newly adapted nomenclature system. Editors of relevant journals will also be informed of the new nomenclature guidelines.

## IV. DISCUSSION

This proposal is intended as a reference guide to aid investigators in properly and consistently naming new HIV-1 strains. It should be emphasized that the working group discussed several different revision models. However, the following considerations were regarded as a priority and thus shaped the ultimate proposal: First, the working group felt that consistency with the existing literature was critically important. Thus, while some participants favoured more extensive changes, the working group ultimately opted for retaining current group and subtype nomenclature. Second, the working group felt that the new nomenclature should have practical applicability. There was consensus that the subtype and group nomenclature has been extremely useful in the past for the tracking of the global AIDS epidemic. As it remains unknown to what extent HIV-1 genetic variation impacts the biological and immunogenecity properties of this virus, molecular epidemiological surveys remain a high priority (this is particularly true for AIDS vaccine research). The new nomenclature proposal has thus been geared toward classifying epidemiologically important viral strains, and is less focussed on naming the incidental viral variant. Third, although all participants recognized that current recombinant nomenclature can be arbitrary, a decision was made to continue to map future recombinants based on current criteria. Tracking more recent (and future) recombination events was considered to be of greater practical value than trying to reconstruct recombination events that might have taken place in the distant past. However, the real possibility that certain presently identified "pure" lineages might represent recombinant forms was certainly recognized. Moreover, it was recognized that the current naming system may in some cases reflect the order in which discoveries were made, rather than the true biological history of the genetic lineages. Finally, it was recognized that many more HIV strains may be recombinant than are currently realized, as recombination may be occurring frequently within subtypes but remains undetected because of insensitive methods.

Although, for the most part, this document represents a consensus view of the working group there was some disagreement among participants at the meeting over the criteria used for assigning recombinants. In particular, the classification of "subtype E" was a point of contention. J.P. Anderson, G.H. Learn, J.I. Mullins, and A.G. Rodrigo disagreed that current evidence favours a recombinant origin of CRF01 (formerly the "subtype E" viruses). These investigators reported that statistical testing using the Kishino-Hasegawa test against the null hypothesis (that no recombination had occurred) did not yield significant results in support of recombination (Anderson *et al.*, submitted). Furthermore, they stressed that certain methods commonly used to infer a recombinant origin of a virus (bootscanning and pairwise distance comparisons) could be indicative of recombination when, in fact, none had occurred (Anderson *et al.* submitted). Thus, JPA, GHL, JIM and AGR argue that there is no conclusive evidence that subtype E has a recombinant origin. While agreeing that formal evidence for a recombinant nature of "subtype E" viruses is lacking (because of the absence of a non-recombinant parental strain), the majority felt that the discordant phylogenetic positions of the different "subtype E" genomic regions are most simply explained as the result of a recombination event. The group felt that a mere acceleration of evolutionary rate in "subtype E" env alone would not be sufficient to move this cluster of viruses outside of subtype A. To support this conclusion, it has been noted that phylogenetic trees constructed from synonymous and non-synonymous changes yielded very similar tree topologies (Gao *et al.* 1996). Thus, the majority of participants agreed that "subtype E" should be designated a circulating recombinant form (CRF01_AE).

As more sequence data have become available, it has become increasingly evident that there is a need to obtain complete genome sequences before a new genetic subtype is established. The best example of this is the former subtype I (now designated "U" within CRF04_cpx) in which four unclassified regions, which were initially thought to all represent the same subtype, were later shown to be derived from at least two different subtypes (Salminen 1999). Although the great majority of HIV-1 recombinants presently characterized are not as complicated as CRF04, this may change in the future as recombinants are likely to encounter, and thus recombine with, other recombinants. This is likely to complicate matters, as it might become impossible to track the many breakpoints and subtype compositions with certainty. Nevertheless, current methods allow the identification of viruses that share the same genome structure, even if they represent complex recombinants, and the revised nomenclature accommodates and emphasizes such strains as they gain a wide-spread geographic distribution.

The "subtypes" and "sub-subtypes" are the distinct clades that are seen in HIV-1 group M phylogenies, but HIV-1 group M viruses have also been characterized that do not group closely to any of the known subtypes. Such viruses have not been designated as representatives of novel subtypes, and are designated under the revised system as unclassified (U). For example, the 1983 Zairian isolate Z3 was shown not to belong to either subtype B or D (Li *et al.* 1988), and still does not cluster with any of the characterized subtypes. As more viruses are isolated, it is likely that more sequences will be found that do not fall neatly into the characterized subtypes. Also, distinct lineages such as B¢ (or Thai B), the subtype B strains circulating in Thailand that form a clade within subtype B (Kalish *et al.* 1995), are not given special consideration because this lineage clearly clusters within subtype B. Thus, it may be unavoidable that this latest nomenclature will require further revision at some point in the future. In the meantime, it is hoped that the current proposal will rectify obvious inconsistencies and provide a reasonable framework for the classification of global strains of HIV-1. Although it is unlikely that there will be a simple relationship between genotype (be it subtype or CRF) and biological phenotype due to the many factors that can select and amplify strains in the pandemic, a clear and consistent genetic classification of strains continues to be useful for epidemiological tracking of the pandemic, vaccine design, and for providing a foundation for detecting biological differences if any do exist.

## Acknowledgments

## References

Anderson, J. P., A. G. Rodrigo, G. H. Learn, A. Madan, C. Delahunty, M. Coon, M. Girard, S. Osmanov, L. Hood, and J. I. Mullins. Testing the hypothesis of a recombinant origin of Human Immunodeficiency Virus Type 1 Subtype E. Submitted.

Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol* **70:**5935-43.

Carr, J. K., B. T. Foley, T. Leitner, M. Salminen, B. Korber, and F. McCutchan. 1998a. Reference sequences representing the principle genetic diversity of HIV-1 in the Pandemic. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences, Los Alamos National Laboratory, Los Alamos, New Mexico.

Carr, J. K., M. O. Salminen, J. Albert, E. Sanders-Buell, D. Gotte, D. L. Birx, and F. E. McCutchan. 1998b. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* **247:**22-31.

Carr, J. K., T. Laukkanen, M. O. Salminen, J. Albert, A. Alaeus, B. Kim, E. Sanders-Buell, D. L. Birx, and F. E. McCutchan. 1999. Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS* **13:**1819-26.

Charneau, P., A.M. Borman, C. Quillent, D. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, F. Clavel. 1994. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* **205**(1):247-53.

Coffin, J. M. 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol* **42:**1-26.

De Leys, R., B. Vanderborght, M. Vanden Haesevelde, L. Heyndrickx, A. van Geel, C. Wauters, R. Bernaerts, E. Saman, P. Nijs, and B. Willems. 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* **64**(3):1207-16.

Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley, and M. P. Busch. 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. *J Virol* **69:**3273-81.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp. 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* **70:**7013-29.

Gao, F., D. L. Robertson, C. D. Carruthers, Y. Li, E. Bailes, L. G. Kostrikis, M. O. Salminen, F. Bibollet-Ruche, M. Peeters, D. D. Ho, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998a. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J Virol* **72:**10234-41.

Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barre-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998b. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J Virol* **72:**5680-98.

Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397:**436-41.

Gurtler, L. G., P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, and L. Kaptue. 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* **68:**1581-5.

Howard, T.M., and S. Rasheed. 1996. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *ARHR* **12:**1413-25.

Hu, W.-S., and H. M. Temin. 1990a. Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *PNAS (USA)* **87:**1556-60.

Hu, W. S., and H. M. Temin. 1990b. Retroviral recombination and reverse transcription. *Science* **250:**1227-33.

Janssens, W., L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J. L. Perret, C. Atende, P. Piot, and G. van der Groen. 1994. Genetic and phylogenetic analysis of env subtypes G and H in central Africa. *ARHR* **10:**877-9.

Kalish, M. L., A. Baldwin, S. Raktham, C. Wasi, C. C. Luo, G. Schochetman, T. D. Mastro, N. Young, S. Vanichseni, H. Rubsamen-Waigmann, H. von Briesen, J.I. Mullins, E. Delwart, B. Herring, J. Esparza, W. L. Heyward, and S. Osmanov. 1995. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* **9:**851-7.

Korber, B. T., S. Osmanov, J. Esparza, and G. Myers. 1994. The World Health Organization Global Programme on AIDS proposal for standardization of HIV sequence nomenclature. *ARHR* **10:**1355-8.

Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J Virol* **69:**6122-30.

Laukkanen, T., J. Albert, K. Liitsola, S. D. Green, J. K. Carr, T. Leitner, F. E. McCutchan, and M. O. Salminen. 1999. Virtually full-length sequences of HIV type 1 subtype J reference strains. *ARHR* **15:**293-7.

Leitner, T., A. Alaeus, S. Marquina, E. Lilja, K. Lidman, and J. Albert. 1995. Yet another subtype of HIV type 1? *ARHR* **11:**995-7.

Leitner, T., S. Kumar, and J. Albert. 1997. Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *J Virol* **71:**4761-70.

Li, W.-H., M. Tanimura, and P. M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Molecular Biology and Evolution* **5:**313-30.

Liitsola, K., I. Tashkinova, T. Laukkanen, G. Korovina, T. Smolskaja, O. Momot, N. Mashkilleyson, S. Chaplinskas, H. Brummer-Korvenkontio, J. Vanhatalo, P. Leinikki, and M. O. Salminen. 1998. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* **12:**1907-19.

Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, G. van der Groen, K. Fransen, G. M. Gershy-Damet, R. Deleys, and D.S. Burke. 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7:**769-80.

Louwagie, J., W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, G. van der Groen, F. E. McCutchan, and D. S. Burke. 1995. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol* **69:**263-71.

McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, and D. S. Burke. 1992. Genetic variants of HIV-1 in Thailand. *ARHR* **8:**1887-95.

McCutchan, F. E., J. K. Carr, M. Bajani, E. Sanders-Buell, T. O. Harry, T. C. Stoeckli, K. E. Robbins, W. Gashau, A. Nasidi, W. Janssens, and M. L. Kalish. 1999. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* **254:**226-34.

Montavon, C., F. Bibollet-Ruche, D. Robertson, B. Koumare, C. Mulanga, E. Esu-Williams, C. Toure, S. Mboup, E. Saman, E. Delaporte, and M. Peeters. 1999. The identification of a complex A/G/I/J recombinant HIV-1 virus in different West African countries. *ARHR* **15:**1707-12.

Murphy, E., B. Korber, M.-C. Georges-Courbot, B. You, A. Pinter, D. Cook, M.-P. Kieny, A. Georges, C. Mathiot, F. Barre-Sinoussi, and M. Girard 1993. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic. *ARHR* **9:**997-1007.

Reviews

Myers, G., S. F. Josephs, A. B. Rabson, T. F. Smith, and F. Wong-Staal. 1988. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, New Mexico.

Myers, G., K. MacInnes, and B. Korber. 1992. The emergence of simian/human immunodeficiency viruses. *ARHR* **8:**373-86.

Nasioulas, G., D. Paraskevis, E. Magiorkinis, M. Theodoridou, and A. Hatzakis. 1999. Molecular analysis of the full-length genome of HIV type 1 subtype I: evidence of A/G/I recombination. *ARHR* **15:**745-58.

Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek. 1994. fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci* **10:**41-8.

Quinones-Mateu, M. E., and E. J. Arts. 1999. Recombination in HIV-1: update and implications. *AIDS Rev* **1:**89-100.

Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. 1995. Recombination in HIV-1. *Nature* **374:**124-6.

Robertson, D. L., and F. Gao. 1998. Recombination of HIV Genomes, Saksena, N. ed, vol. Human Immunodeficiency Viruses. Medical Systems SpA, Rome.

Roques, P., E. Menu, R. Narwa, G. Scarlatti, E. Tresoldi, F. Damond, P. Mauclere, D. Dormont, G. Chaouat, F. Simon, and F. Barre-Sinoussi. 1999. An unusual HIV type 1 env sequence embedded in a mosaic virus from Cameroon: identification of a new env clade. *ARHR* **15:**1585-9.

Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, and A. Mayer. 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* **68:**6340-6.

Salminen, M. O., C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995a. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* **213:**80-6.

Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995b. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *ARHR* **11:**1423-25.

Salminen, M. O., B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. Leinikki, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence of an ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. *ARHR* **12:**1329-39.

Salminen, M. 1999. Unpublished observations.

Simon, F., P. Mauclere, P. Roques, I. Loussert-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi, and F. Brun-Vezinet. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* **4:**1032-7.

Triques, K., A. Bourgeois, S. Saragosti, N. Vidal, E. Mpoudi-Ngole, N. Nzilambi, C. Apetrei, M. Ekwalanga, E. Delaporte, and M. Peeters. 1999. High diversity of HIV-1 subtype F strains in Central Africa. *Virology* **259:**99-109.

Triques, K., A. Bourgeois, N. Vidal, E. Mpoudi-Ngole, C. Mulanga-Kabeya, N. Nzilambi, N. Torimiro, E. Saman, E. Delaporte, and M. Peeters. 2000. Near-full-length genome sequencing of divergent African HIV-1 subtype F viruses leads to the identification of a new HIV-1 subtype designated K. *ARHR* **16:**139-151.

Reviews