Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated ‘subtype I’, reveals a unique complex A/G/H/K/? mosaic pattern

D. Paraskevis,1, M. Magiorkinis,1 A. M. Vandamme,2 L. G. Kostrikis3 and A. Hatzakis1

1 National Retrovirus Reference Center, Department of Hygiene and Epidemiology, Athens University Medical School, M. Asias 75, Athens 115 27, Greece
2 Rega Institute for Medical Research and University Hospitals, Katholieke Universiteit Leuven, Belgium
3 Aaron Diamond AIDS Research Center, The Rockefeller University, New York 10016, USA

Human immunodeficiency virus type 1 (HIV-1) has been classified into three main groups and 11 distinct subtypes. Moreover, several circulating recombinant forms (CRFs) of HIV-1 have been recently documented to have spread widely causing extensive HIV-1 epidemics. A subtype, initially designated I (CRF04_cpx), was documented in Cyprus and Greece and was found to comprise regions of sequence derived from subtypes A and G as well as regions of unclassified sequence. Re-analysis of the three full-length CRF04_cpx sequences that were available revealed a mosaic genomic organization of unique complexity comprising regions of sequence from at least five distinct subtypes, A, G, H, K and unclassified regions. These strains account for approximately 2% of the total HIV-1-infected population in Greece, thus providing evidence of the great capability of HIV-1 to recombine and produce highly divergent strains which can be spread successfully through different infection routes.

According to the updated proposed nomenclature system, human immunodeficiency virus type 1 (HIV-1) has been divided into groups, subtypes, sub-subtypes and circulating recombinant forms (CRFs) (Robertson et al., 2000). CRFs represent intersubtype recombinant viruses that have been spread successfully and share an identical pattern of mosaicism (Robertson et al., 2000). For the designation of a new CRF, representative strains must be identified in at least three individuals with no direct epidemiological linkage (Robertson et al., 2000; http://hiv-web.lanl.gov/). There are currently four CRFs: (a) CRF01_AE, representing AE recombinants originating from Central Africa and having been extensively spread in Asia (Carr et al., 1996; Gao et al., 1996); (b) CRF02_AG, representing AG recombinants circulating in West and Central Africa (Carr et al., 1998, 1999); (c) CRF03_AB, representing AB recombinants, which originated and spread in Kaliningrad (Bobkov et al., 1998; Liitsola et al., 1998); and (d) CRF04_cpx, representing recombinant viruses found in Cyprus and Greece that were initially characterized as A/G/I (Gao et al., 1998; Nasioulas et al., 1999) but which were recently found to be a more complex recombination of genomic sequences derived from subtypes A, G, H, K as well as regions of unclassified sequence (Robertson et al., 2000). To date, Cyprus and Greece are the only geographical regions where these CRFs have been detected. These strains account for approximately 2% of the total HIV-1-infected individuals in Greece (Paraskevis et al., 1999a).

This study presents a re-analysis of the three full-length sequences of CRF04_cpx available from Cyprus and Greece (strains 94CY032.3, 97PVMY, 97PVCH) and includes all known HIV-1 subtypes A–D, F–H, J and K.

To determine any potential relationships of these strains with subtypes in addition to A and G, we performed a bootscanning plot using the Ray Simplot software, version 2.5 (http://www.med.jhu.edu/deptmed/stray/download/). All three available full-length strains of CRF04_cpx were each separately, and subsequently together, compared with reference isolates of subtypes A–D, F–H, J and K (Fig. 1). More specifically, trees were constructed for a window of 400 bp moving in steps of 50 bp and then the obtained bootstrap values supporting the clustering of the CRF04_cpx isolates with the different subtypes were plotted across the alignment (Gao et al., 1996). The separate bootscanning plots showed similar patterns, suggesting that the three strains can indeed be considered a CRF. The bootscanning plot suggested that CRF04_cpx isolates consisted of genomic regions belonging to subtypes A, G, H and K as well as regions which showed no apparent similarity to these subtypes (Fig. 1). Subtype...
classification for each region was confirmed by phylogenetic analysis by the neighbour-joining method with Kimura’s two-parameter correction (Kimura, 1980) using programs from the PHYLIP package, version 3.5 (Felsenstein, 1993). The reliability of the phylogenetic trees produced was estimated by bootstrapping (100 replicates). Only representative trees are shown in Fig. 2. More specifically, clustering of CRF04_cpx sequences in the middle of the pol gene, in the region spanning nt 2000–3000, was significantly favoured for both subtypes F and K (Fig. 1). Using this region, CRF04_cpx isolates cluster on a deep branch defined by both subtypes F and K, but are more closely associated with subtype K (Fig. 2). The clustering of subtype G in the region spanning nt 5251–5650 was not supported with high bootstrap values in the bootscanning plot (Fig. 1). Phylogenetic analysis in this region revealed that CRF04_cpx sequences were similar to those of subtypes G and J, but were only distantly related to these subtypes (Fig. 2).

On the basis of these results, the suggested recombination pattern of the full-length sequences of CRF04_cpx is shown in Fig. 3. According to this scheme, CRF04_cpx presents an extremely complex pattern of mosaicism, including at least 15 breakpoints that occurred between subtypes A, G, H, K and the parental HIV-1 isolate from which the unclassified regions originated. More specifically, the p17/p24 gag region, the 3’ pol/vif region, the first exons of tat and rev and the 5′ terminus of gp41 classified as subtype A. The p7/p6 gag region, the pol region and the 5′ portion of gp120 clustered with subtype G and the 3’ region of gp41 as well as a middle portion of pol and the 3’ region of gp120 fell within subtypes H and K, respectively. On the other hand, the remaining genomic regions could not be classified with any of the previously characterized HIV-1 subtypes and thus remained unclassified. These regions will be classified if non-recombinant ‘parental’ subtype I sequences become available.

Furthermore, we re-examined the subtype classification of three other recombinant HIV-1 strains MAL, Z321 and BFP90, which were found to cluster with the unclassified portions of CRF04_cpx in partial genomic regions (Alizon et al., 1986; Getchell et al., 1987; Srinivasan et al., 1989; Gao et al., 1998; Montavon et al., 1999). A bootscanning plot was performed for the complete pol gene of MAL and BFP90 isolates as well as for the genomic region spanning pol, vif/vpr and tat/vpu of the Z321 sequence, with a sliding window of 400 bp moving in steps of 50 bp. According to the boot-
Fig. 2. Representative phylogenetic trees of different genomic regions. Positions in the alignments are shown above each tree. The number at the nodes indicates bootstrap values. Only values greater than 75 are shown. 94CY032.3, 97PVMY and 97PVCH sequences are boxed.
Fig. 3. (a) Mosaic structure of CRF04_cpx genome organization. Partial regions named as subtypes A, G, H, K and unclassified regions are shown in red, green, yellow, pink and white, respectively. Regions with uncertain classification are shown in grey.

(b) Phylogenetic trees of BFP90 (nt 2070–3120) and MAL (nt 2070–2970) where both sequences clustered with subtype K, and Z321 (nt 4372–4711) where it clustered with CRF04_cpx sequences.
scanning plot, MAL and BF90 showed a high degree of similarity to subtype K in these regions, which was initially classified as ‘subtype I’. Phylogenetic analysis further confirmed these observations, as shown in Fig. 3. Interestingly, the above regions coincided with the partial pol region of CRF04_cpx classified as subtype K (Fig. 3). On the other hand, the Z321 isolate was found to cluster with the CRF04_cpx sequences in partial vif/vpr regions as initially described (Fig. 3) (Gao et al., 1998). According to these results, Z321 is the only recombinant to date that has been found to cluster with the unclassified regions of the CRF04_cpx in partial genomic regions.

Re-analysis of the three available full-length sequences of CRF04_cpx suggested that the three isolates share a common pattern of complex mosaicism consisting of at least five distinct HIV-1 subtypes. The observed recombination pattern was quite similar to that presented in the HIV-1 nomenclature proposal (Robertson et al., 2000). More specifically, 3’ pol/vif and 5’ env are derived from subtype A and G, respectively, and not unclassified regions. To date, CRF04_cpx are unique in mosaicism complexity, representing the only recombinant HIV-1 strains consisting of more than four different HIV-1 subtypes which have spread successfully in several cases (Paraskevis et al., 1999a, b). More specifically, the 94CY032.3 strain initially designated as ‘subtype I’ in 1995 was first identified in Cyprus in two heterosexual partners who presumably became infected in Greece (Kostrikis et al., 1995; Nasioulas et al., 1998). Until then, CRF04_cpx strains had been detected in several cases in Greece, accounting for approximately 2% of HIV-1-infected individuals (Paraskevis et al., 1999a). According to epidemiological data, these strains have not been restricted to particular routes of HIV-1 transmission, as initially suggested (Nasioulas et al., 1999), but were also found in different risk group categories such as heterosexuals, vertically infected children, intravenous drug users, as well as in a homosexual male (Paraskevis et al., 1999a). Until now, CRF04_cpx sequences have been detected only in Cyprus and Greece, but they possibly originated from Africa, for two reasons. (a) CRF04_cpx isolates present a complex pattern of mosaicism consisting of subtypes A, G, H, K and U. These subtypes, and especially subtypes H and K, are very uncommon in geographical regions outside Africa, thus suggesting that the putative recombination event between these subtypes possibly occurred in Africa. (b) A partial genomic region clustering with the unclassified portion of CRF04_cpx has been documented for one other recombinant, Z321, from Africa (Getchell et al., 1987; Srinivasan et al., 1989; Gao et al., 1998).

The unique complexity of the CRF04_cpx HIV-1 strains provides evidence for the great capability of the HIV-1 genome to recombine and to give rise to new viruses with an altered genetic make-up. The identification of these recombinant strains in several cases of HIV-1-infected individuals in Greece, infected through different routes, suggests the successful spread of these variants. However, the question that still remains unanswered is whether these strains have altered biological properties associated with their high genetic complexity.

We wish to acknowledge Giikas Magiorkinis for producing the figures. This study was supported by the Hellenic Society for the Study of AIDS and Sexually Transmitted Diseases. D.P. was supported by the Hellenic Center for Infectious Diseases. L.G.K. was supported by grants partly provided by the National Institutes of Health under ROI AI43868.

References


Received 17 July 2000; Accepted 25 October 2000