Evolution of the HIV-1 V3 region in the Italian epidemic

Luigi Buonaguro, Maria Tagliamonte, Marialina Tornesello, Franco M. Buonaguro
Laboratory of Viral Oncogenesis and Immunotherapy and AIDS Reference Center, National Cancer Institute "Fond. G. Pascale", Naples, Italy

The epidemic of Human Immunodeficiency Virus type 1 infection in Italy is mostly ascribed to the B subtype, which represents the prevalent subtype in Western Countries. The virus isolates of the B subtype, moreover, show an increasing nucleotide heterogeneity over time, indicating a continuous intra-subtype dynamic evolution, typical of long-lasting epidemics. In recent years, however, the progressive decrease in the transmission rate among the historically defined risk groups (i.e. homosexuals and IDUs) and the parallel increase in heterosexual transmission are slowly introducing variants of non-B subtype into the Italian HIV-1 epidemic. This appears to be strictly linked to the growing number of immigrants from non-Western Countries, where non-B clades and CRFs are prevalent, and consequent inter-racial blending.

The distribution of these novel genetic forms needs to be evaluated by continuous molecular monitoring nationwide to verify whether they will overcome the pre-existing B-clade epidemic, which could have significant implications for diagnosis, treatment and vaccine development.

Here we review the genetic evolution of HIV-1 spreading within the Italian epidemic.

KEY WORDS: HIV-1, V3 region, Italy, Molecular evolution

GLOBAL HIV-1 GENETIC VARIATION

HIV-1 genetic subtypes
The Human Immunodeficiency Virus type 1 (HIV-1) is characterized by an extensive genetic heterogeneity generated by the lack of proof-reading ability of the reverse transcriptase (RT) (Roberts, J.D. et al., 1988), the rapid turnover of HIV-1 in vivo (Ho, D. D. et al., 1995), host selective immune pressures (Michael, N.L., 1999) and recombination events during replication (Temin, H.M., 1993). This variability has led to the development of a subtype nomenclature for the classification of virus isolates, with the designation of a group M (main), a group O (outlier) and a group N (non-M/non-O) (Ayoub, A. et al., 2000; Gurtler, L. et al., 1994; Simon, F. et al., 1998). Group M, responsible for the majority of infections in the HIV-1 epidemic worldwide, can be further divided into 10 recognized phylogenetic subtypes or clades (A-K), which are approximately equidistant from one another (Figure 1).

Within group M, the average inter-subtype genetic variability is 15% for the gag protein, and 25% for the env protein (Janssens, W. et al., 1994; Kostrikis, L.G. et al., 1995; Kuiken, C. et al., 2000; Leitner, T. et al., 1995a; Myers, G. et al., 1992; Robertson, D.L. et al., 2000a).

Moreover, within a subtype it is possible to identify groups of isolates forming genetically related sister clades, termed sub-subtype (Robertson,
D.L. et al., 2000a), which appear to be phylogenetically more closely related to each other than to other subtypes. This is the case of the A and F clades, whose members are currently classified in A1 - A2 and F1 - F2 sub-subtypes, respectively (Gao, F. et al., 2001; Triques, K. et al., 2000). B and D clades are also more closely related to each other than to other subtypes, but their original designation as subtypes is retained by the authors for consistency with earlier published work.

The original HIV-1 subtype classification was based on sub-genomic regions of individual genes. With the increasing number of viral isolates available worldwide and the improvement in sequencing methods, HIV-1 phylogenetic classifications are currently based either on nucleotide sequences derived from multiple subgenomic regions (gag, pol and env) of the same isolates or on full-length genome sequence analysis. This approach has revealed virus iso-

**FIGURE 1** - Evolutionary relationships among non-recombinant HIV-1 strains. The phylogenetic tree shows the M (main), O (outlier) and N (Non-M/Non-O) HIV-1 groups and, within the M group, the subtypes. The phylogenetic analysis was performed on near-full length sequences and based on the neighbor joining method. The reliability of the internal branches defining a subtype was estimated from 100 bootstrap replicates.

**FIGURE 2** - Mosaic structure of HIV-1 circulating recombinant forms (CRFs). The CRFs described to date are shown (modified from http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html). Letters and graphic patterns represent the different subtypes of HIV-1 involved in the recombination events; U stands for “Unknown”.

L. Buonaguro, M. Tagliamonte, M.L. Tornesello, F.M. Buonaguro
lates in which phylogenetic relations with different subtypes switch along their genomes. These inter-subtype recombinant forms are thought to have originated in individuals multiply infected with viruses of two or more subtypes. When an identical recombinant virus is identified in at least three epidemiologically unlinked people, and is characterized by full-length genome sequencing, it can be designated as a circulating recombinant form (CRFs) (Peeters, M., 2001; Robertson, D. L. et al., 2000b). More than 20 CRFs have been reported to date, whose origin can be tracked in areas where the parental strains are co-circulating (Figure 2). The importance of CRFs in the global HIV-1 pandemic is increasingly recognized, accounting for 18% of incident infections (Osmanov, S. et al., 2002) and representing the local predominant form in Southeast Asia (CRF01-AE) (Menu, E. et al., 1996; Motomura, K. et al., 2000; Piyasirisilp, S. et al., 2000) or in West and West-Central Africa (CRF02-AG) (McCutchan, F. E. et al., 1999; Montavon, C. et al., 2000). The co-circulation of multiple subtypes and CRFs in the same population is increasing the probability of individuals “superinfected” with different HIV-1 genetic forms which can swap parts of their genetic material. This is resulting in the generation of several recombinants called “unique recombinant forms,” or URFs, which, if spread to other people, will lead to “circulating recombinant forms” (CRFs) (McCutchan, F.E., 2006).

**Geographic distribution of HIV-1 subtypes**

Molecular epidemiological studies show that, with the exception of Sub-Saharan Africa, where almost all subtypes, CRFs and several URFs are detected, there is a specific geographic distribution pattern of HIV-1 subtypes (Osmanov, S. et al., 2002) (Figure 3). This seems to be the consequence of either accidental trafficking (viral migration), with a resulting “founder” effect, or a prevalent route of transmission, resulting in a strong advantage and local predominance of earlier HIV-1 subtypes over subsequently introduced genetic forms (Buonaguro, L. et al., 1995;

On a global scale, according to recent studies, the most prevalent HIV-1 genetic forms are subtypes A, B, C and CRF02 AG, where the subtype C accounts for almost 50% of all HIV-1 infections worldwide. Most infections in countries in the southern part of Africa, India, and Ethiopia are caused by subtype C, which is also circulating as a minor form in Brazil and Russia. Subtype A viruses are predominant in areas of central and east Africa (Kenya, Uganda, Tanzania, and Rwanda), in east European countries formerly constituting the Soviet Union and in Middle East Countries, where they are mainly transmitted among injecting drug users (Naderi, H.R. et al., 2006). Subtype B is the main genetic form in western and central Europe, the Americas, and Australia, and is also common in several countries of southeast Asia, northern Africa, the Middle East, and among South African and Russian homosexual men. The recombinant virus CRF02 AG is the most prevalent genetic form throughout all of west Africa and in parts of central Africa (Osmanov, S. et al., 2002).

HIV-1 EPIDEMIC IN ITALY

Modification in the pattern of transmission

The first phase of the HIV epidemic in Italy was mainly confined to the injecting drug users (IDUs) risk group, with an absolute predominance of HIV-1 B subtype, in accordance with other Western Countries. In particular, considering the total AIDS cases reported in the adult population during the period between 1982 and 2004, 61.0% were IDUs (including also homosexual IDUs) with similar percentages in men and women (65.2% and 58.7%, respectively) (Italian AIDS Operative Center, 2005). However, the annual percentages of AIDS cases reported in IDUs have gradually decreased from 69.7% in 1987 to 32.3% in 2004 (Suligoi, B. et al., 2004), in part as a consequence of prevention programs implemented in Italy to discourage syringe sharing (Nicolosi, A. et al., 1991; Rezza, G. et al., 1993). In parallel, the AIDS cases reported in heterosexual individuals account for 19.5% of total epidemic cases, with a significantly higher percentage in women compared to men (41.2% vs 13.6%). However, the annual percentage of AIDS cases related to heterosexual transmission has dramatically increased over the years, becoming the most prevalent risk factor for AIDS in 2004 (40.4%) (Figure 4A) (Italian AIDS Operative Center, 2005).

Although almost 25% of heterosexual individuals diagnosed with AIDS in Italy are partners of long-term HIV-1 infected individuals, carrying a “historical” B-subtype virus, more than 10% of them are either immigrants from regions endemic for HIV-1 (6.87%) or their Italian partners (3.03%). Moreover, the risk is not identified for 64% of them (Figure 4B). This epidemiological evidence, based only on the AIDS reported cases and not considering all the HIV-1 infections derived also from traveling abroad, suggests that at least 10% of the viruses transmitted through heterosexual contacts could potentially belong to non-B subtypes and CRFs. This has been recently reported in other European Countries, with a higher prevalence due to an older tradition of immigration waves and much tighter historical as well as economic links with Countries endemic for HIV-1 infection (Boni, J. et al., 1999; Couturier, E. et al., 2000; Devereux, H. et al., 1999; Fransen, K. et al., 1996; Holguin, A. et al., 2002; Iversen, A. K.N. et al., 1999; Leitner, T. et al., 1995b; Paraskevis, D. et al., 1999; Snoeck, J. et al., 2002).

HIV-1 sub-genomic sequences reported in Italy

During the epidemic in Italy, several HIV-1 subgenomic nucleotide sequences have been reported and deposited in the Los Alamos Database, although the first “near full-length” sequences were reported and deposited only very recently (Tagliamonte, M. et al., 2006). Considering the three structural gag, pol and env genes, only sequences of the env gene have been collected during a long period of the Italian epidemic, starting from the mid 80’s (Op de Coul, E.L. M. et al., 2001), and can be used for a comprehensive evaluation of the HIV-1 genetic evolution in the Italian epidemic (Bagnarelli, P. et al., 2004; Bagnarelli, P. et al., 1999; Balotta, C. et al., 1997; Binley, J. M. et al., 2004; Buonaguro, L. et al., 1994; Buonaguro, L. et al., 2004; Buonaguro, L. et al., 2002; Casado, C. et al., 2000; Halapi, E. et al., 1997; Mammano, F. et al., 1995; Menzo, S. et al., 1998; Salvatori, F. et al., 1997;
On the contrary, pol sequences (the protease region, in particular) have been extensively analyzed and collected only from the year 2000, when the ART-induced resistant viral isolates started to appear in heavily drug-treated HIV-1-infected individuals and in recent seropositive individuals, naïve for drug treatment (Balotta, C. et al., 2000; Balotta, C. et al., 2001; Ceccherini-Silberstein, F. et al., 2004; Cinque, P. et al., 2001; La Seta-Catamancio, S. et al., 2001; Monno, L. et al., 2003; Perno, C. F. et al., 2001; Romano, L. et al., 2000; Saracino, A. et al., 2006; Svicher, V. et al., 2006; Venturi, G. et al., 2002). Finally, the only gag sequences in individuals resident in Italy, immigrants from regions endemic for HIV-1 or their Italian partners, were recently identified in our studies (Buonaguro, L. et al., 2004; Buonaguro, L. et al., 2002; Tagliamonte, M. et al., 2006).

HIV-1 GENETIC EVOLUTION IN THE ITALIAN EPIDEMIC

Continuous evolution in the B-clade C2-V3 env sequences in the Italian epidemic

Phylogenetic analysis has been performed including all the B-subtype Italian sequences described over the entire HIV-1 epidemic in Italy, covering the C2-V3 env region. The resulting phylogenetic pattern shows an “Italian branch” clearly separated from B clade US and Thai sequences.
used as the “B-clade outgroup”. Moreover, within the Italian branch, the HIV-1 isolates are distributed in three major clusters, each of them including several sub-clusters (Figure 5). The sequences derived from the different studies do not form independent clusters and/or sub-clusters but are found inter-dispersed in the tree. This result is not surprising considering that the majority of these reports are based on samples identified in Italy mainly during overlapping periods covering the first half of the 90’s. Likewise, a specific distribution pattern of the sequences based on the risk factor for HIV-1 infection (IVDU, Homo- or Heterosexuality) are not observed.

The vast majority of the sequences are found in the B1 and B2 clusters and, in particular, the B1 cluster includes the majority of sequences identified over a broad range of time. The B3 cluster, instead, is prevalently (7/13, 53.8%) but not solely based on recent sequences identified in our study.

Phylogenetic analysis, therefore, strongly suggests that the B subtype HIV-1 epidemic in Italy can be derived from three main molecular ancestors, which have continuously evolved and spread among infected individuals during the epidemic regardless of the prevalent transmission route.

**Non-B-clade C2-V3 env sequences in the Italian epidemic**

Throughout the HIV-1 epidemic in Italy, only seven non-B clade env sequences have been described: two G, two CRF-02AG and one A (Buonaguro, L. *et al.*, 2004; Buonaguro, L. *et al.*, 2002); one A1B (Bagnarelli, P. *et al.*, 2004); one CRF-12BF (Binley, J.M. *et al.*, 2004). In particular, the five non-B subtype sequences described by our group were identified in heterosexual individuals either immigrants from sub-Saharan Africa or their Italian partners (Buonaguro, L. *et al.*, 2004; Buonaguro, L. *et al.*, 2002). The non-B isolate described by Bagnarelli *et al.* was identified in a Ugandan child, stably resident in Italy, infected by mother-to-child transmission. Instead, for the case reported by Binley *et al.* there is a lack of information on the transmission route.
and ethnicity of the infected individual. In addition, a very recent near-full length sequence analysis has shown that the HIV-1 isolates, originally classified as A and G, are actually divergent strains not clustering in any of the known subtypes. In particular, the A isolate is close to the A3 sub-subtype but potentially represents a novel sub-subtype to be confirmed with the identification of at least two additional related isolates in unlinked individuals (Tagliamonte, M. et al., 2006) (Figure 6).

**Preferential codon usage in the C2-V3 env sequences in the Italian epidemic**

The env sequences were subsequently analyzed for codon usage in the octamer located at the tip of the V3 loop of the env gene (GPGRAFYT). In particular, the usage of a G or a C (as silent mutation) in the third position of the GGN codon encoding the second glycine residue ($G_3$), was previously associated with HIV-1 isolates identified in patients with different risk practices. In fact, the GGG codon has been associated with the homosexual risk group and the GGC codon with the IDU risk group (Adwan, G. et al., 1999; Casado, C. et al., 2000; Kuiken, C. et al., 1994; Kuiken, C. et al., 1996). Our analysis shows that, considering the Italian B subtype sequences, the GGC codon is found in 56.7% of the variants isolated from IVU individuals, in 26.7% of those isolated from heterosexual and 3.5% of those isolated from homosexual individuals. On the contrary, the GGG codon is found in 10.8% of the variants isolated from IDU individuals, in 66.7% of the variants isolated from heterosexual and 89.3% of those isolated from homosexual individuals ($p<0.0001$) (Figure 7). These results confirm that the GGC codon is strongly associated with i.v. transmission and the GGG codon with sexual (homo and hetero) transmission of HIV-1. The variants identified in the mother-to-child group show the GGG codon in 32.4% and the GGC codon in 62.9% of cases, respectively, suggesting i.v. as well as sexual transmission in this heterogeneous group. The other two silent codons GGA and GGT, on the contrary, are found at low percentages (10.8% – 18.5%) with no significant differences among the risk groups ($p = 0.027$). The segregation of the GGC and GGG codons in the variants transmitted through the i.v. and the sexual route could be the outcome of either viral tropism, genetic bottlenecks or a founder effect and may represent a possible fingerprint pattern for predicting the transmission route.

**CONCLUDING REMARKS**

The analysis of the env gene’s C2-V3 region of the HIV-1 isolates identified during the Italian epidemic shows that the B clade is still predominant and is circulating among all risk groups. The B subtype HIV-1 variants show an extensive genetic diversification and the currently circulating viruses appear to derive from at least three ‘founders’ which arose during the earlier phases of the Italian epidemic. On the contrary, the introduction and spread of non-B subtype HIV-1 isolates in the Italian epidemic appear to be relatively recent and still limited, but strong-
ly associated with heterosexual transmission. Moreover, the individuals infected by non-B clade isolates are either immigrants from sub-Saharan Africa or their Italian partners, as reported in other European Countries. This would also explain the lower prevalence of non-B clades in Italy compared to other European Countries with an older tradition of immigration waves and much tighter historical and economic links with African Countries. This observation, based on env sequences, is in accordance with what has been reported in Italian studies performed on the pol gene of HIV-1 variants identified in the most recent years of the Italian epidemic.

As consequence of the results obtained on selected cohorts, the distribution of multiple HIV-1 clades and recombinant forms needs to be evaluated on a larger scale by coordinated nationwide molecular monitoring. It is extremely important to establish whether the new genetic forms will overcome the pre-existing B-clade epidemic and whether the B and non-B subtype segregation in different risk groups will persist during the epidemic. These events should be monitored for the possible future development of adequate diagnostic, treatment and prevention strategies.

ACKNOWLEDGEMENTS
This study was supported by grants from the Ministero Italiano della Sanità (Ricerca Corrente and Progetto Finalizzato AIDS 2002) and the ICSC-World Lab, Lausanne, Switzerland (Project MCD-2/7).

REFERENCES


Buonaguro, L., Tagliamonte, M., Tornesello, M. L.,...


A recent outbreak of human immunodeficiency virus type 1 infection in Southern China was initiated by two homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC. J. Virol. 74, 11286-11295.


