Running title: Near full-length sequence analysis of HIV-1 BF recombinants from Italy

Francesco Saladini[1], Brian T. Foley[2], Andrea Rosi[1], Ilaria Vicenti[1], Giulio Nannetti[1], Genny Meini[1], Francesca Razzolini[1], and Maurizio Zazzi[1]

1 Section of Microbiology, Department of Biotechnology, University of Siena, Siena, Italy.
2 HIV Databases, T-6, Mail Sop K710, Los Alamos National Lab, Los Alamos, NM 87545 USA

Corresponding author:
Francesco Saladini,
Department of Biotechnology, Section of Microbiology
University of Siena
Viale Bracci, 1
53100 Siena, Italy
Phone +39 0577 233910
Fax +39 0577 233870
Email: saladini6@unisi.it

Abstract

Recombination between HIV-1 subtypes B and F has generated several circulating and unique recombinant forms, particularly in Latin American areas. In Italy, subtype B is highly prevalent while subtype F is the most common pure non-B subtype. To investigate the recombination pattern in Italian BF recombinant viruses, we characterized full-length sequences derived from 15 adult patients, mostly Italian and infected by heterosexual route. One of the BF mosaics was a CRF29, three sequences clustered with low bootstrap values with CRF39, CRF40 and CRF42. With the exception of the CRF29-like sequence, the other recombination patterns were unique but two possible clusters were identified. Analysis of the gp120 V3 domain suggested a possible link with subtype F from Eastern Europe rather than from Latin America, favoring the hypothesis of local recombination between clade B and F viruses over that of import of BF recombinants from Latin America. HIV-1 subtypes B and F appear prone to generation of unique recombinants in Italy, warranting epidemiological surveillance and investigation of a possible clinical significance.

Human Immunodeficiency Virus type 1 (HIV-1) is characterized by an extremely broad genetic diversity, caused by high mutation and recombination rates. Currently, HIV-1 group M ("Major"), which is responsible of the worldwide pandemic, is divided into pure subtypes A-D, F-H, J, K (with subtypes A and F further subdivided into sub-subtypes A1-A4 and F1-F2, respectively) as well as 48 Circulating Recombinant Forms (CRFs) of varying epidemiological significance, and untold numbers of unique recombinants which are not known to have spread to many patients. Subtype B has been long predominant in Western European countries but the prevalence of non-B strains has been increasing significantly in latest years as a consequence of recent migration waves from Africa, Eastern Europe and South America[1]. Continuous surveillance of HIV-1 molecular epidemiology plays a critical role in the understanding of genetic diversity of HIV-1 and for research purposes such as vaccine development. In clinical practice, pol-based subtype assignment is usually accomplished as a by-product of genotypic antiretroviral resistance testing.

A homebrew HIV-1 genotyping assay has been established and offered as a public health service at the HIV Monitoring Laboratory (HML), Department of Molecular Biology, University of Siena, Italy since 1995[2]. The laboratory has been serving a number of clinics for analysis of drug resistance mutations accumulating around 10,000 protease and reverse transcriptase sequences from more than 4,000 patients. In addition, some hundred partial gag and env sequences have been obtained for research studies. A recent survey on the whole HIV-1 sequence database at the HML revealed 85% of subtype B sequences and 15% of non-B subtypes, mainly CRF02_AG, F1, C and A1, based on the pol region[3]. In some cases, assignment of subtype based on pol and env regions were not in agreement[4]. These discrepancies mostly involved subtypes B and F1. In addition, other sequences showed BF1 mosaics within the individual pol or env region. Since BF recombinants have been previously reported in Latin America[5-11], we were interested in analyzing BF mosaic forms circulating in Italy.

Samples from 15 adult patients (12 males, three females) collected in four hospitals of Tuscany and Umbria regions were selected for near full-length sequencing based on peripheral blood mononuclear cell (PBMC) DNA availability and identification of a BF pol sequence or discordant BF pol and env. Twelve patients were Italian, two were Brazilian patients living in Italy and 1 was a Tunisian patient; most of the patients acquired HIV-1 infection via sexual intercourses, only one patient contracted HIV-1 infection through drug injection (Table 1). To obtain full-length sequence information, PBMC DNA was subjected to a 9.010-base pair pre-amplification step[12] followed by multiple nested PCRs generating overlapping sub-genomic fragments. The LTR region, not comprised in the large outer amplicon, was obtained separately using outer primers MZ28 (coordinates 60-79 in HIV-1 HXB2) and LR56 (1488-1507) and inner primers P221 (87-100) and P82 (1478-1505). PCR products were sequenced on both strands using multiple infrared-labeled primers and sequencing products were resolved with a Licor IR2 dual-laser automated sequencer. PCR and sequencing primer sequences are available on request. Nucleotide sequences were edited and assembled by the DNAstar SeqMan II module. For phylogenetic analysis, full-length sequences were compared with full-length subtype reference sequence alignments obtained from the Los Alamos National Laboratory HIV Sequence Database (http://www.hiv.lanl.gov/). The alignment contained nucleotide sequences of all HIV-1 group M subtypes and known CRFs with BF mosaics. The HIV-1 group O sequence SEMP1300 was introduced in the

alignment as the outgroup for the phylogenetic tree. The sequence dataset was manually managed using the BioEdit program, version 5.0.9 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html/) (T. Hall, North Carolina State University, Raleigh, NC). Phylogenetic analysis was performed with PAUP4b10 software. Neighbor joining and maximum likelihood trees were constructed based on the fittest nucleotide substitution models determined by Modeltest v3.7. The reliability of the tree topology was assessed by bootstrapping with 1,000 replicates. The recombination patterns were determined by bootscanning analysis with Simplot v3.5.1 (http://sray.med.som.jhmi.edu/SCRoftware/simplot/), using a window of 400 nt, moving in 50 nt increments, which made it possible to assign segments to subtypes B and F1 with a bootstrap support ≥70%. Prediction of coreceptor use was done by using the geno2pheno[coreceptor] algorithm at the 10% false positive rate (http://coreceptor.bioinf.mpi-inf.mpg.de/).

Phylogenetic analysis of the near full-length sequences is shown in Fig. 1. Sample 59211 was identified as a CRF29, while samples 83166, 30638 and 85037 clustered with low bootstrap values with CRF39, CRF40 and CRF42, respectively. Bootscanning analysis showed the high genetic diversity between reference sequences and our BF mosaics, which are mostly characterized by unique recombination patterns. Two distinct clusters of BF sequences shared a similar recombination pattern, as shown in Fig. 2. Each BF cluster may have originated from a common ancestor circulating among people living in the same geographical area. Further recombination events may then have introduced different breakpoints contributing to diversification within the same cluster. This hypothesis is supported by greater sequence similarity among members of cluster 1 and among members of cluster 2, than between clusters or the unique recombinants. Cluster 2 showed a high degree of relatedness and may represent a new circulating recombinant form if complete genomes of other isolates are completed in the future.

BF recombinants where first isolated and characterized in Argentina, where they have been circulating since mid-80's[5], and later in Brazil[6]. The presence of these recombinants and related unique mosaic strains increased over the years, favored by the high number of injecting drug users, sexual contacts and the cocirculation of subtype B, F and BF viruses in the same populations[7,8]. Afterwards, BF mosaics were described in other Southern America countries[9,10] as well as in European countries like Spain[13] and Luxembourg (where CRF42_BF has been isolated). Recently, a full-length sequence of a BF recombinant strain with a unique mosaic pattern has been characterized in Italy[14] in a patient with multiple risk factors and found to be phylogenetically related to two other sequences from Brazil. Most of the BF genomes described in this study were acquired via heterosexual contacts. BF mosaics could have been imported into Italy from Southern America or locally generated by independent recombination events between subtype B, which is predominant in Italy, and subtype F1, which is the most frequent non-B pure subtype circulating in Italy. The migration flow from South America to Italy and sexual tourism in Brazil could support the first hypothesis. Although most of the sequences reported here do not seem to be related to the CRFs originated in Brazil or Argentina and other BF sequences found in GenBank (data not shown), phylogenetic analysis of the pol region indicated that of the seven BF recombinants with an F1 pol are more similar to the F1 subtype circulating in South America than to that circulating in Eastern Europe (supplementary Figure 1).

Analysis of the gp120 V3 loop sequence predicted 11 CCR5-tropic and 4 CXCR4-tropic viruses. The V3 crown tetrapeptide was GPGR in 11/15 (73.3%) sequences, GPGQ in 2/15 (13.3%), GPGG and GPGK in the other two cases. Among 9 clade B V3 sequences, GPGR motif was observed in 7 cases, while the other 2 sequences had GPGG and GPGK motifs. Among 6 clade F1 V3 sequences, GPGR was found in 4 cases and GPGQ in the other 2 samples. None of the samples with subtype B V3 region had the GWGR motif, which is common in clade B viruses circulating in Brazil but not in other countries[15,16]. Interestingly, two of the subtype F1 V3 sequences had the GPGQ crown variant, which is largely prevalent in Romania and is not present in Brazilian F1 (mostly GPGR). Thus, this last finding could support the hypothesis of the local origin of BF recombinants due to the co-circulation of the subtype B and F1.

The appearance of these mosaic structures in Italy can be compared to that occurred in Argentina, where subtype B was predominant before the introduction of subtype F1. Actually, the prevalence of BF recombinants has been increasing in Latin American countries, becoming the most common genetic form and causing the disappearance of pure subtype F in some areas[5,15]. The high propensity for subtype B and F to recombine could be explained by a higher fitness compared with pure subtype B and F1[11]. To test this hypothesis, extensive in vitro analysis of chimeric viruses is required to investigate the interaction among subgenomic regions from each subtype[17]. In general, the prevalence of BF, as well as other, mosaic structures is probably underestimated due to large scale partial genotyping of clinical samples for antiretroviral resistance testing. Along with upcoming availability of effective technologies for large scale sequencing, information on full-length HIV genomes is expected to expand providing an accurate picture of virus diversification in different areas over time.

Sequence data

GenBank accession numbers of the sequences are GU595148 - GU595162.

Acknowledgements

This work was supported by grants from the Italian Ministry of Health (PRIN grant n. 200887SYZ5 and AIDS Program grant n. 40h81) and from the European Community under the Seventh Framework Program (CHAIN project, grant n. 223131).

References

1. Paraskevis D, Pybus O, Magiorkinis G, et al.: Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology* 2009; 6:49.
2. Peduzzi C, Pierotti P, Venturi G, Romano L, Mazzotta F, Zazzi M. Performance of an in-house genotypic antiretroviral resistance assay in patients pretreated with multiple human immunodeficiency virus type 1 protease and reverse transcriptase inhibitors. J Clin Virol 2002; 25:57-62.
3. Lai A, Riva C, Marconi A, et al.: Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. *HIV Med.* 2010; 11:593-602.

4. Razzolini F, Vicenti I, Saladini F, et al.: Natural variability in the HR-1 and HR-2 domains of HIV type 1 gp41 from different clades circulating in Italy. *AIDS Res Hum Retroviruses* 2007; 23:558-563.
5. Gomez Carrillo M, Avila M, Hierholzer J, et al.: Mother-to-child HIV type 1 transmission in Argentina: BF recombinants have predominated in infected children since the mid-1980s. *AIDS Res Hum Retroviruses* 2002; 18:477-483.
6. Sabino EC, Shpaer EG, Morgado MG, et al.: Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 1994; 68: 6340-6346.
7. Thomson MM, Delgado E, Herrero I, et al.: Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J Gen Virol* 2002; 83:107-119.
8. Sanabani S, Kleine Neto W, Kalmar EM, Diaz RS, Janini LM, Sabino EC: Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in São Paulo, Brazil. *Infect Genet Evol* 2006; 6:368-377.
9. Carr JK, Avila M, Gomez Carrillo M, et al.: Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS* 2001; 15:F41-F47.
10. Sierra M., Thomson MM, Rios M, et al.: The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. *Infect Genet Evol 2005*; 5:209-17.
11. Aulicino PC, Holmes EC, Rocco C, Mangano A, Sen L: Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina. *J Virol* 2006; 81:427-429.
12. Salminen MO, Koch C, Sanders-Buell E, et al.: Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* 1995; 213: 80-86.
13. Fernández-García A, Pérez-Alvarez L, Cuevas MT, et al.: Identification of a new HIV type 1 circulating BF intersubtype recombinant form (CRF47_BF) in Spain. *AIDS Res Hum Retroviruses* 2010; 26:827-32.
14. Bruselles A, Rozera G, Bartolini B, et al.: Use of massive parallel pyrosquencing for near full-length characterization of a unique HIV type 1 BF recombinant associated with a fatal primary infection. *AIDS Res Hum Retroviruses* 2009; 25:937-942.
15. Morgado MG, Sabino EC, Shpaer EG, et al.: V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses* 1994; 10:569-576.
16. Leal E, Villanova FE.: Diversity of HIV-1 subtype B: implications to the origin of BF recombinants. *PLoS One.* 2010; 5:e11833.
17. Carobene MG, Rodrìgues CR, De Candia CA, Turk G and Salomon H: In vitro dynamics of HIV-1 BF recombinants genomic regions involved in the regulation of gene expression. *Virol J* 2009; 6:107.

Address reprint requests to:
Francesco Saladini,

Department of Biotechnology, Section of Microbiology
University of Siena
Viale Bracci, 1
53100 Siena, Italy
Phone +39 0577 233910
Fax +39 0577 233870
Email: saladini6@unisi.it

Table 1. Epidemiological data of the patients included in this study.

Patient sample	Year of HIV diagnosis	Year of sample collection	Sex	Nationality	Mode of transmission
30638	1990	1999	M	Italian	Bisexual
61842	1994	2002	F	Italian	Heterosexual
79867	1994	2005	M	Brazilian	?
57954	1995	2002	M	Italian	Heterosexual
59211	1998	2002	M	Italian	Heterosexual
89072	1998	2006	M	Brazilian	?
85037	1999	2005	M	Italian	Homosexual
50610**	2001	2001	F	Italian	?
50612**	2001	2001	M	Italian	?
53143*	2001	2001	M	Italian	Heterosexual
58736*	2002	2002	M	Italian	Heterosexual
59352**	2002	2002	F	Italian	Heterosexual
62136*	2002	2002	M	Italian	Heterosexual
83166	2005	2005	M	Italian	?
104045	2007	2007	M	Tunisian	Intravenous drug user

*Cluster 1 sequences and **Cluster 2 sequences based on phylogenetic tree as shown in figure 1.