Migration patterns of HIV-1 subtype B virus in Northern Italy

Alessandra Lo Presti1, Massimo Ciccozzi1, Eleonora Cella1, Marta Giovanetti1, Gianguglielmo Zehender2, Daniela Valentì3, Nathalie Iannotti3, Emanuele Malfatto3, Claudia Bernardini3, Franco Maggiolo3, Annapaola Callegaro4

1 Department of Infectious, Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy; 2 Department of Clinical Sciences L Sacco, University of Milan, Italy; 3 Division of Infectious Diseases, Ospedali Riuniti di Bergamo, Italy; 4 Microbiology and Virology, Ospedali Riuniti di Bergamo, Italy

Summary

Gene flow analysis is used to identify the migration patterns of viruses within a geographical area and/or in different populations. 883 HIV-1 B subtype pol gene sequences were analyzed. The gene analysis among different geographical areas of the Bergamo district and from different transmission risk groups showed 25% of the observed gene flow was from people living in the north valleys to lowland and 40.5% from a heterosexual risk group to injecting drug users. Injecting drug users seem to be the central link, mercenary sex being the common route of transmission (and gene flow) between this group and both heterosexual and homosexual individuals.

KEY WORDS: HIV-1 subtype B epidemic, Phylogenetic analysis, Migration patterns analysis.

Received June 7, 2012
Accepted September 6, 2012

Current epidemiological reports in Italy show that at the end of the year 2010 62,617 individuals were affected by AIDS and in the period 1985-2009 a total of 45,707 new HIV-1 diagnoses were reported by the regional/provincial systems (Suligoi et al., 2010). Overall, it is estimated that 140,000/150,000 HIV-positive people are living in Italy. Regarding the route of transmission for HIV-1, sexual contacts (both heterosexual and homosexual) have become the leading cause of infection (79%) (Suligoi et al., 2010). Cases are not homogeneously distributed throughout the nation and the region of Lombardy in the Northern part of the country contributes with 18,574 AIDS cases to the general picture. Within the region, the highly populated and industrialized province of Bergamo shows one of the highest national incidence rates of new cases: 3.6 cases per 100,000 inhabitants (Suligoi et al., 2010). Although non-B-subtype infections have been reported in Italy (Baldanti et al., 2008; Balotta et al., 2001; Buonaguro et al., 2008; Ciccozzi et al., 2007), the HIV-1 epidemic among long-term residents is still mainly sustained by subtype B viruses. Recently, through a phylogenetic approach, we identified epidemiological networks and sequence interrelationships between acute/early and chronic HIV-1 type B infections in both drug-naive and drug-experienced individuals representing almost all infected individuals in Northern Italy (Callegaro et al., 2011).

The district of Bergamo is located in Eastern Lombardy and occupies the central section of the Lombardy Pre-Alps and a small part of the Po plain, with a population of 1,027,162 inhabitants. A map of the territory shows that mountains, hills and the plain occupy, from North to South, three wide horizontal bands. Heterogeneous characteristics of this highly populated region can contribute to a different spatial diffusion and migration history of HIV-1 viruses.
The migration flow of viral lineages may provide information on HIV epidemics and recent studies investigated by in/out flow migration approach the spread of HIV-1 subtypes in Europe (Salemi et al., 2008a; Salemi et al., 2008b). Here, we evaluated a large dataset of subtype B HIV-1 virus sequences sampled from different areas of the Bergamo province to attempt to define migration events of infection across this a local, well-limited area. These insights could direct prevention strategies as well as the management of treatment programs within the territory.

A total of 883 HIV-1 subtype B pol gene sequences (Protease [PR] and the first 344 reverse transcriptase [RT]) codons were sampled from 305 drug-naive and 578 drug-treated HIV-1 patients to represent almost all the infected individuals living in the Northern part of Italy. All sequences were collected through routine drug resistance testing between 2000 and 2008 at the Division of Infectious Diseases of Bergamo, the district reference site as already described (Callegaro et al., 2011).

Patients were from all the geographically distinct areas in which the Bergamo province was divided, considering the heterogeneous characteristics (mountains, hills, plains) and the highly urbanized areas of the Bergamo territory (Figure 1a). The sampling location of all the isolates were valley (n=170), west area (n=50), lake east (n=137), lowland (n=336) and city and hinterland (n=106); 84 sequences were from people living out of the Bergamo province but followed up by Bergamo Division of Infectious Diseases. The prevalence of the transmission risk group among the study population is reported in Table 1.

![Figure 1a](image1a.png)

**FIGURE 1 a, b - a) The sampling location of all HIV-1 B isolates. V, Valley (n=170), WA, West area (n=50), LE, Lake East (n=137), LL, Lowland (n=336), CH, City and Hinterland (n=106), OD, outside the province (n=84). Source: “Official website of the Bergamo’s District SITer®”. The arrows indicate the gene flow to/from different geographic areas of the district of Bergamo and the thickness of the arrows is proportional to the amount of gene flow (migrations obtained with Mac Clade analysis). b) Maximum parsimony migration patterns of subtype B HIV-1 sequences to/from different areas of Bergamo province. The bubblegram shows the frequency of gene flow (migrations) to/from different geographic areas, as the percentage of the total observed migrations estimated from the maximum likelihood tree with a modified version of the Slatkin and Maddison test. The surface of each circle is proportional to the percentage of observed migrations given within the circle. V, Valley; WA, West area; LE, Lake East; LL, Lowland; OD, Outside the province; CH, City and Hinterland."
The first 1302 nucleotides of pol gene from 883 HIV-1 B subtype infected patients were analyzed. All the sequences were aligned, removing gaps and cutting to identical sequence lengths as already described (Callegaro et al., 2011). A maximum likelihood phylogenetic tree was generated with the GTR + I + G model of nucleotide substitution. The evolutionary model was chosen as the best-fitting nucleotide substitution model in accordance with the results of the hierarchical likelihood ratio test (HLRT) as already described (Ciccozzi et al., 2011). The statistical robustness and reliability of the branching order within the phylogenetic tree was confirmed by the approximate likelihood-ratio test (aLRT) approach; all calculations were performed using Phyml v 3.0 (Guindon et al., 2010). This tree was used as starting tree to perform the gene flow analysis.

The hypothesis of restricted gene flow among distinct HIV-1 B populations in geographically distinct areas of Bergamo province was tested with a modified version of the Slatkin and Maddison test (Slatkin and Maddison, 1989) to counts HIV-1 B migrations, using the Mac-Clade version 4 program (Sinauer Associates, Sunderland, MA). A one-character data matrix was obtained from the original data set by assigning to each taxon in the tree a one-letter code indicating its own sampling location. Then, the putative origin of each ancestral sequence (i.e., internal node) in the tree is inferred by finding the most parsimonious reconstruction (MPR) of the ancestral character. The final tree length, i.e., the number of observed migrations in the genealogy, can easily be computed and compared to the tree-length distribution of 10,000 trees obtained by random joining-splitting. Observed genealogies significantly shorter than random trees indicate the presence of subdivided populations with restricted gene flow. Moreover, similar gene flow analysis was also performed counting the HIV-1 B exchanges to/from different transmission risk groups. A one-character data matrix was obtained from the original data set by assigning to each taxon in the tree a one-letter code assigning at each sequence its transmission risk group. Migrations analysis among different risk groups (character states) were traced with the State changes and stasis tool (MacClade software), as described above. The maximum likelihood phylogenetic tree built with Phyml was analyzed with a modified version of the Slatkin and Maddison method (Slatkin and Maddison, 1989) to infer the HIV-1 B gene flow (migration) among different areas of Bergamo province. To perform this analysis, the HIV-1 sequences were assigned to six distinct groups: V, valley; WA, west area; LE, lake east; LL, lowland; OD, outside the province; CH, city and hinterland. The null hypothesis of panmixia (i.e., no population subdivision or complete intermixing of sequences from different geographic areas) was rejected by the randomization test (p<0.0001) (Slatkin and Maddison, 1989). Most of the observed HIV-1 B gene flow (25%) was from valley to low land (Figure 1a, b). A gene flow

| Table 1 - Prevalence of the transmission risk group among study population. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Risk Group                      | MSM N. (%)      | IDUs N. (%)     | Heterosex N. (%)| Bisex N. (%)    | Others N. (%)   | Total           |
| Valley                          | 32 (18.8)       | 54 (31.8)       | 80 (47)         | 1 (0.6)         | 3 (1.8)         | 170             |
| West area                       | 10 (20)         | 20 (40)         | 19 (38)         | 1 (2)           |                 | 50              |
| Lake East                       | 15 (11)         | 63 (46)         | 58 (42.3)       | 1 (0.7)         |                 | 137             |
| Lowland                         | 43 (12.8)       | 170 (50.6)      | 118 (35.1)      | 3 (0.9)         | 2 (0.6)         | 336             |
| City and hinterland             | 36 (34)         | 44 (41.5)       | 26 (24.5)       |                 |                 | 106             |
| Out of district                 | 18 (21.4)       | 34 (40.5)       | 29 (34.5)       | 2 (2.4)         | 1 (1.2)         | 84              |
| Total                           | 154             | 385             | 330             | 6               | 8               | 883             |
was also observed from outside the province to
the valley area (9.4%), while a small percentage of
HIV gene flow was observed from lake east to
outside the province (4.5%) and from lake east to
city and hinterland (3.4%) (Figure 1a, b). On the
contrary, no gene flow was observed from valley
to west region and from west region to valley. The
lowland, city and hinterland showed no migra-
tion flow to any other province area.
The different risk factors were also analyzed in
terms of gene flow. This analysis was performed
using the same phylogenetic tree previously de-
scribed, to infer the HIV-1 B gene flow among
different risk groups. To perform this analysis,
the HIV-1 sequences were assigned to four risk
groups: hom, homosexual; di, injecting drug users; he, heterosexual; bi, bisex-ual. Most of the observed HIV gene flow (40.5%) was from the heterosexual risk group to injecting drug users
(Figure 2).
As far as gene flow/risk factors for HIV are con-
cerned, 18.7% of the observed HIV gene flow was
from injecting drug users to homosexual indi-
viduals. Small HIV gene exchanges (respectively
6.6% and 3.9%) were also observed from homo-
sexuals to injecting drug users and from homo-
sexual individuals to bisexuals (Figure 2). On the
contrary, no gene flow was observed from ho-
mosexual to heterosexual and vice versa.
We report, for the first time in Italy, the migra-
tion patterns of HIV-1 B subtype lineages by gene
flow analysis, counting viral migrations to/from
different geographical areas of a heterogenous
and highly populated Northern Italian province.
The migration analysis was performed on a large
cohort of infected individuals, in which we pre-
viously demonstrated a cluster distribution within
the observed epidemics (Callegaro et al., 2011).
The heterogenous characteristics of this region, in
terms of population distribution, can contribute
to a different spatial diffusion of HIV-1 viruses
and one could expect that the highest populated
areas such as the city, hinterland and low land,
could be the major source of infection.
Unexpectedly most of the observed HIV-1 B gene
flow (25%) was from people living in the north
valleys to individuals resident in the industrial-
ized and highly populated low land. Furthermore,
the valley showed a significant infl ow migration
(9.4%) from people living outside Bergamo
province.
This pattern is difficult to explain. A possible ap-
proach would consider socio-economic charac-
teristics. People living in the industrialized low-
land or in the city/hinterland area usually work
and spend their free time within the same area.
On the contrary, a much higher migration was
observed from the mountain region.
Assigning HIV subtype B sequences to four risk
groups, the different risk factors were also ana-
yzed in terms of gene flow. Most of the HIV gene
flow (40.5%) was observed from the heterosexual
risk group to injecting drug users. 18.7% of the
observed HIV gene flow were from injecting drug
users to homosexual individuals. This pattern is
also difficult to analyze, however a possible ap-
proach should consider the dynamics of the dif-
f erent risk groups. Both injecting drug users and
homosexuals are quite “closed” entities whose be-
haviour usually remains within the risk group.
on the other hand heterosexuals can mix widely.
It can be hypothesized that injecting drug users
are the central link, mercenary sex being the com-
mon route of transmission (and gene flow) be-
tween this group and both heterosexual and ho-
mosexual individuals.
These insights could direct prevention strategies as well as the management of treatment programs in a local, specific and well-characterized territory. As an example, since people in the valley area the most common source of virus, it could be cost effective to concentrate in this area any campaign aimed to inform people about the prevention of sexually transmitted diseases.

REFERENCES


