Infection, Genetics and Evolution 14 (2013) 232–239

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Phylogenetic and demographic characterization of HIV-1 transmission in Madrid, Spain ☆

Gonzalo Yebra [a], África Holguín [a,*], Deenan Pillay [b], Stéphane Hué [b]

[a] HIV-1 Molecular Epidemiology Laboratory, Department of Microbiology, Hospital Ramón y Cajal-IRYCIS and CIBERESP, Ctra. Colmenar Viejo, Km 9.100, 28034 Madrid, Spain

[b] Division of Infection and Immunity, Medical Research Council Centre for Medical Molecular Virology, University College London, Cruciform Building, Gower St., WC1E 6BT London, United Kingdom

ARTICLE INFO

Article history:
Received 22 October 2012
Received in revised form 10 December 2012
Accepted 12 December 2012
Available online 2 January 2013

Keywords:
HIV/AIDS
Transmission networks
Phylogenetic analysis
Bayesian MCMC
Madrid
Spain

ABSTRACT

Background: The combination of phylogenetic analyses of HIV sequences with patients' demographic data allows us to understand local HIV transmission, a necessary knowledge for designing prevention strategies. The Community of Madrid represents a challenge for the control of HIV epidemic in Spain given its high HIV prevalence and increasing proportion of immigrant people among HIV-infected population.

Methods: We applied maximum likelihood methods and Bayesian Markov chain Monte Carlo (MCMC) inference using the program BEAST to a set of HIV-1 *pol* sequences from 1293 patients diagnosed in 1995–2010 in Madrid, Spain.

Results: Two-hundred and thirty six patients (18.2% of the cohort) were included in 100 transmission chains using phylogenetic criteria, 67 (67%) belonging to HIV-1 subtype B and 33 (33%) to 11 different non-B strains, especially BG and BF recombinants. Most networks involved transmission between MSM (48/100). Half of non-B clusters (15/33) included at least one Spaniard. Sub-Saharan African patients presented a low linkage rate (9%) in contrast to Spanish (21%) and Latin American (25%) patients. Three clusters involving treatment-independent transmission of drug-resistance mutations were found.

Conclusions: One out of five HIV-infected patients in our cohort in Madrid was epidemically linked, mainly by transmission pairs. The inclusion in transmission networks was more likely for MSM, Spaniards and patients from Latin America. We found no evidence of self-sustained non-B epidemics due to the absence of large transmission chains with the exception of Cuban BG recombinants and CRF47_BF. However, the differences in transmission across variants are probably determined by the patient profile, especially the infection route.

1. Introduction

By 2009, 130,000 people lived with HIV/AIDS in Spain and the estimated HIV prevalence among adult population was 0.4% (ECDC, 2010; UNAIDS, 2011). This is one of the highest HIV prevalences in Western Europe. Nevertheless, the incidence of newly diagnosed HIV infections has stabilized around 8 infections per 100,000 people per year. HIV in Spain is transmitted mainly through unprotected sexual intercourse, especially among men having sex with men (MSM), who accounted for 42.5% of new infections in 2009 (Díez et al., 2012; ECDC, 2010).

As a consequence of the significant increase in the proportion of legal immigrant population in Spain (from 1.6% in 1998 to 12.2% in 2010) (INE, 2011), 37.6% of the HIV newly diagnosed patients in 2009 were people coming from other countries (Díez et al., 2012; ECDC, 2010), mainly from Latin America (19.4%) and sub-Saharan Africa (9.4%). The Community of Madrid, containing the capital and the most populous city in the country –Madrid–, is a key region in the Spanish HIV/AIDS epidemics: it has accounted for 20% of all the AIDS cases ever detected in Spain and 47.2% of the HIV newly diagnosed in the region in 2009 were foreign-born (Díez et al., 2012) due to the constant arrival of people from other countries.

One of the implications of immigration has been the introduction of HIV-1 strains from developing regions. These imported viruses tend to be of HIV subtypes other than B, while subtype B viruses predominate in resource-rich countries. However, HIV-1 non-B variants are now rising in these countries due to their spread

☆ Parts of the data were presented at the 10th European Meeting on HIV & Hepatitis – Treatment Strategies & Antiviral Drug Resistance, Barcelona (Spain), March 2012; and at the 4th GESIDA National Congress, Toledo (Spain), November 2012.

* Corresponding author. Address: HIV-1 Molecular Epidemiology Laboratory, Department of Microbiology, Hospital Universitario Ramón y Cajal, Ctra. Colmenar Viejo, Km 9.100, 28034 Madrid, Spain. Tel.: +34 91 3368330; fax: +34 91 3368809.
E-mail address: aholguin.hciii@salud.madrid.org (Á. Holguín).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.12.006

in both immigrant and autochthonous population (Brennan et al., 2010; Hemelaar et al., 2011; Vercauteren et al., 2009). They account for 10–15% of new infections in Spain, including mostly recombinant forms (González-Alba et al., 2011; Holguín et al., 2008; Yebra et al., 2012).

The increasing accessibility of HIV-1 genetic sequences has permitted phylogenetic methods to recreate the history of HIV-1 epidemics. Partial HIV-1 *pol* gene sequences routinely obtained for genotypic drug resistance testing are adequate to infer transmission events and to characterize epidemiological patterns of public health relevance (Hué et al., 2004).

Phylogenetic analyses have been used to understand the contribution of recent HIV infections to onward transmission (Brenner et al., 2007; Pao et al., 2005), drug-resistance transmission in untreated populations (Brenner et al., 2008; Callegaro et al., 2011; Hué et al., 2009; Skoura et al., 2011; Yerly et al., 2009), HIV transmission in specific risk groups like MSM (Fisher et al., 2010; Lewis et al., 2008; Zehender et al., 2010), heterosexual population (Hughes et al., 2009) or injecting drug users (Skar et al., 2011), and the introduction and spread of different HIV-1 subtypes in Western countries (Gifford et al., 2007; von Wyl et al., 2011). Specifically, Bayesian Markov chain-Monte Carlo (MCMC) approaches that include molecular clock models have greatly contributed by placing HIV-1 transmissions in a real time-frame, allowing the estimation of the origin and tempo of HIV spread (Brenner et al., 2011; Callegaro et al., 2011; Chalmet et al., 2010; Gifford et al., 2007; Hué et al., 2009; Hué et al., 2005; Hughes et al., 2009; Lewis et al., 2008; Zehender et al., 2010). Here, we describe HIV-1 transmission networks involving patients attending HIV/AIDS centers in the Community of Madrid, Spain, using phylogenetic means, showing that one out of five HIV-infected patients in our cohort was epidemically linked. MSM, Spaniards and Latin Americans were more frequently included in transmission networks than patients of other risk group or origin. We observed different non-B sub-epidemics coming from sub-Saharan Africa (CRF02_AG) and Latin America (BF and BG recombinants).

2. Materials and methods

2.1. Cohort description

We collected 1293 previously published (Holguín et al., 2008; Yebra et al., 2010; Yebra et al., 2011) HIV-1 partial *pol* sequences from different patients attending HIV/AIDS clinics in Madrid, Spain, during 1995–2010: Ramón y Cajal Hospital (n = 237; 18.3%), Carlos III Hospital (n = 322; 24.9%), 12 de Octubre Hospital (n = 71; 5.5%) and Sandoval Health Center (n = 213; 16.5%). Gender, reported route of HIV infection, country of origin and treatment status of the patients were recorded when available, as well as the specimens sampling date (month, year). Sequences were sampled in the following periods: 1995–1999 (n = 47; 3.6%), 2000–2003 (n = 323; 25%), 2004–2005 (n = 330; 25.5%), 2006–2007 (n = 435; 33.6%) and 2008–2010 (n = 158; 12.2%). Overall, 798 (61.7%) patients were naïve to antiretroviral treatment at sampling time. In most cases (1189, 92%) the sequence included the complete protease (PR) and partial reverse transcriptase (RT) with a mean length of 1130nt. In 104 cases (8%), only the PR sequence (297 nt) was available.

Among the patients with known transmission mode (n = 831), 373 (44.9%) were heterosexuals, 326 (39.2%) were MSM, 114 (13.7%) were injecting drug users (IDU) and 18 (2.2%) reported other risk practices. For 462 subjects (35%), the route of infection was unavailable. When known (79.8%), a total of 722 (70%) patients were male and 310 (30%) were female.

The patients were native Spaniards (n = 613, 47.4%), sub-Saharan Africans (n = 300, 23.2%), Latin Americans (n = 217, 16.8%), West Europeans or North Americans (n = 49, 3.8%), Eastern Europeans or Russian (n = 36, 2.8%) and from other regions or unknown (n = 78, 6%).

2.2. HIV-1 subtyping

HIV-1 subtypes and circulating recombinant forms (CRF) were identified by phylogenetic analysis of the *pol* sequences. Representative sequences of the 9 subtypes and the 49 CRF of HIV-1 group M known to date were downloaded from Los Alamos HIV sequence database (http://www.hiv.lanl.gov) and used as references. Sequences were aligned using ClustalX 2.0.11 (Thompson et al., 1997). The phylogenetic tree was obtained using the neighbor-joining method under the Kimura two-parameter model of nucleotide substitution, using the PHYLIP software package (Felsenstein, 1989). A bootstrap analysis (1000 replicates) was performed to test the tree's robustness, setting a cut-off at 700 to classify HIV variants. Unique recombinant variants (URF) were detected using SimPlot v3.5.1 (http://sray.med.som.jhmi.edu/SCRoftware/simplot/), and jumping profile Hidden Markov Model (jpHMM, http://jphmm.gobics.de/jphmm.html).

Sequences belonged to subtype A (n = 49, 3.8%), B (n = 814, 62.9%), C (n = 29, 2.2%), D (n = 8, 0.6%), F (n = 26, 2%), G (n = 53, 4.1%), H (n = 7, 0.5%), recombinant CRF02_AG (n = 177, 13.7%) and other subtypes and recombinants (n = 130, 10%). For the subsequent analyses, we grouped closely related CRFs sharing very similar recombination patterns in recombinant families following the suggestion of (Zhang et al., 2010). Specifically, we grouped the Cuban BG (CRF20, CRF23 and CRF24) and the BF originated in South America (CRF12 and CRF17) in two families.

2.3. Drug resistance

The presence of transmitted drug resistance was defined according to the drug resistance mutations (DRM) list recommended by the World Health Organization (Bennett et al., 2009) using the Calibrated Population Resistance tool (http://cpr.stanford.edu).

2.4. Detection of transmission clusters

Three datasets were created: (i) subtype B sequences (n = 814), (ii) recombinant CRF02_AG (n = 177) and (iii) non-B, non-CRF02_AG variants (n = 302). To detect local sequences not included in the sampling but potentially involved in transmission clusters, the uncorrected pairwise genetic distances (p-distances) were calculated using PAUP* v4.0beta10 (Swofford, 1998). Every sequence that differed by less than 0.045 nucleotide substitutions per site (4.5%) from at least one another sequence in the alignment was selected (Lewis et al., 2008). In order to identify additional HIV-1 sequences involved in the putative transmission networks, the 10 most similar samples for each sequence was retrieved using BLAST (http://blast.ncbi.nlm.nih.gov). In the posterior analyses, those clusters 'broken down' by the addition of foreign sequences were excluded from the analysis. After removing duplicates, a total of 2003 sequences (1043 from our cohort and 960 from public databases) were then used for maximum likelihood (ML) phylogenetic reconstruction under the general time reversible model of nucleotide substitutions with gamma distributed rate heterogeneity (GTR + Γ), using RAxML v7.2.8-alpha (Stamatakis et al., 2008) for the non-B variants and FastTree v2.1.3 (Price et al., 2010) for subtype B due to the higher number of sequences. The robustness of the trees' topologies was assessed by bootstrap analysis (1000 replicates) and by Shimodaira-Hasegawa (SH)-like local support