THE COLOMBIAN EPIDEMIC IS DOMINATED BY HIV-1 SUBTYPE B: A MOLECULAR EPIDEMIOLOGY AND PHYLODYNAMIC STUDY

Andrea-Clemencia Pineda[1,2], Nuno Rodrigues Faria[1], Francisco-Javier Diaz[3], Patricia Olaya[4], Casper Møller Frederiksen[5], Li Guangdi[1], Arley Gomez-Lopez[2], Philippe Lemey[1] , Anne-Mieke Vandamme[1,6]

E-mail: andreapinedap@gmail.com, annemie.vandamme@uz.kuleuven.ac.be

[1]Clinical and Epidemiological Virology, Rega Institute for Medical Research, Department of Microbiology and Immunology, University of Leuven, Leuven, Belgium , [2] Clinical and Molecular Infectious Diseases Group, Universidad del Rosario, Bogotá, Colombia, [3] Inmunovirology group, Universidad de Antioquia, Medellín, Colombia, [4] Centro de Análisis Molecular, Bogotá, Colombia, [5] University of Copenhagen, Faculty of Health Sciences, Copenhagen, Denmark, [6] Centro de Malária e Outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal

Introduction

Although HIV-1 subtype B predominates in Latin America, the proportion of recombinant forms has doubled in the last decade (Hemelaar et al, 2011). Colombia was considered one of the countries with most migration in the region due to economical reasons and armed conflict (OIM Colombia, 2010). In addition, this country has had ~0.5% of HIV prevalence in the last years, which is the highest prevalence in the Andean region (UNAIDS, 2012). However, little is known about the viral diversity and phylodynamics. Two studies, one from Medellin and one from Bogota showed the predominance of subtype B between 1995 and 2002 (Montano et al, 2005; Sanchez et al, 2006), and one study reported a patient infected by subtype F in 2002 (Eyzaguirre et al, 2002), suggesting that non-B subtypes should be further evaluated. We aimed to shed light on the molecular epidemiology and spatial dynamics of the HIV-1 epidemic in Colombia between 2002 and 2007.

Methodology

Nucleotide data. 610 *pol* sequences from patients of 7 Colombian geographic regions sampled during 2002-2007 were obtained from the *Centro de Análisis Molecular,* a center that has performed the majority of resistance testing nationwide since 2000. Sequences were generated by population sequencing using TRUGENE® kit (Siemens Healthcare Diagnostics, Germany).

Sequence selection. For each Colombian sequence, the 10 most similar sequences BLASTn were selected (Yebra et al, 2013). All sequences were subtyped using REGA version 3 (Pineda-Peña et al, 2013) and COMET Version 2 (Struck et al, 2010). In case of discordant classification, manual phylogenetic analysis was used. Colombian sequences and the BLASTn sequences were aligned and edited by using SeaView (Gouy et al, 2010). 37 codon positions associated with drug-resistance mutations were removed (Bennet et al, 2009). The best-fitting nucleotide substitution model was chosen using jModeltest (Posada et al., 2008) and a maximum likelihood tree (ML) was performed with 1000 bootstrap replicates.

Phylogeography. To investigate the spatial diffusion of subtype B in Colombia, we performed phylogeographic analyses (Lemey et al., 2009) using a discrete asymmetric diffusion model implemented in BEAST (Drummond and Rambaut, 2007). We applied a Bayesian Stochastic Search Variable Selection (BSSVS) procedure to estimate the most significant pathways of viral dispersal within Colombian and between several other regions as defined by the BLASTn outcome. Moreover, we quantified viral migration among locations using a robust counting procedure (O'Brien et al., 2009, Minin and Suchard, 2008).

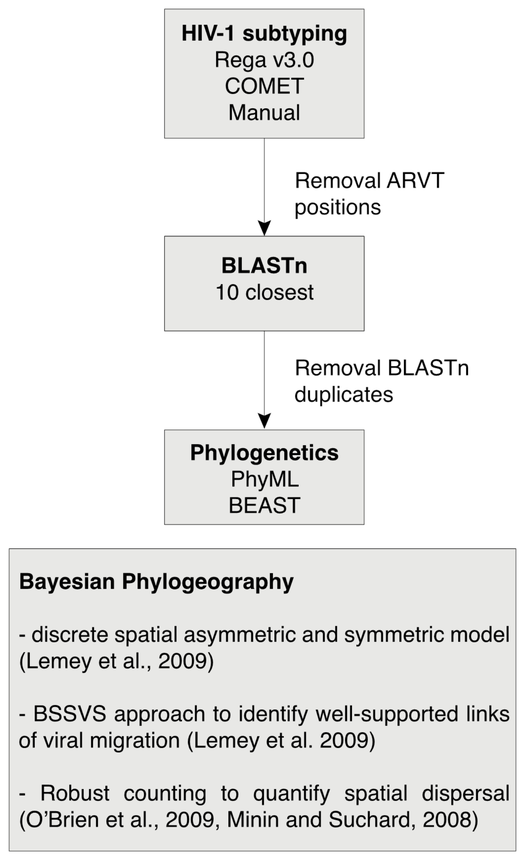
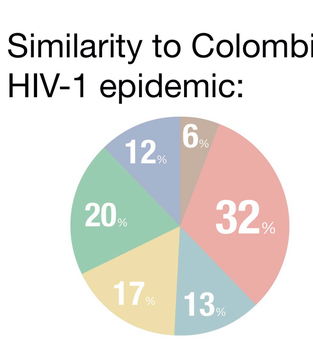


Figure 1. Workflow of the approach followed in this study.

Results and Discussion

The Colombian epidemic remains dominated by subtype B (609/610, 99.8%) in all geographical regions. One subtype F (1/610, 0.2%) was found in Bogotá in 2003. The most similar sequences retrieved by BLASTn for the Colombian subtype F were from Italy, and for subtype B were mainly from Spain (see Figure 2).



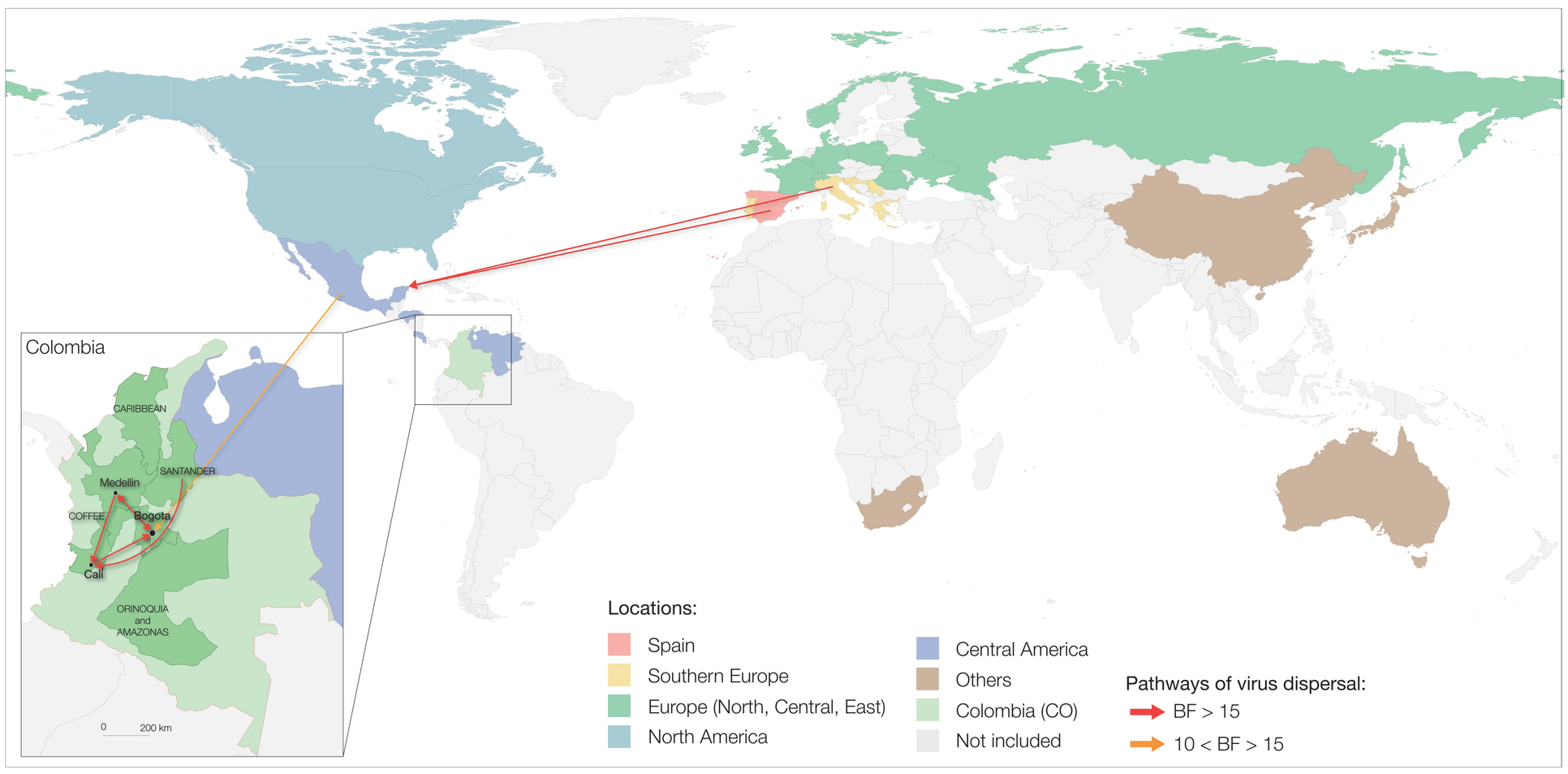


Figure 2. Most similar sequences retrieved by BLASTn (above) and epidemiological links with the Colombian Epidemic (right). Number of sequences per Colombian location was roughly proportional to HIV-1 prevalence as reported in UNGASS 2010. Sequences were collected in Bogotá (275), Cali (77), Medellin (64), Coffee region (42), Santander (26), Caribbean (18), Unknown (108). Unknown locations were estimated using the Bayesian phylogeographic framework using an ambiguity code. BF: Bayes Factor links obtained by the BSSVS procedure.

We found evidence for a mean of 11 introductions of subtype B in Colombia-Bogotá from Central America. In addition, we estimated 34 introductions from Spain to Central America. Within Colombia, we found support for bi-directional flow between Bogotá and Medellin. The robust counting analysis further indicated that Cali and Bogotá were both the main exporters and importers of HIV-1 dispersal at the country-level, suggesting that these two locations play an active role in the countrywide dispersal of HIV-1.

Conclusion

- This is the first molecular epidemiology study of HIV-1 in different geographic regions of Colombia. Our findings show that HIV-1 epidemic in Colombia is greatly dominated in all considered locations by subtype B. Future studies including more recent data should be performed to confirm these findings.
- We show that viral populations in Bogotá and Cali play a central role in shaping the HIV-1 epidemic within Colombia, which has most likely been initially seeded from Central American countries. Moreover, our findings suggest that Spain – and not other Central American or South American locations - was likely the origin of the Central American HIV-1 subtype B that gave origin to the Colombian epidemic.
- Overall, this study provides a simple and flexible framework for future phylodynamic studies targeted at investigating the HIV-1 dynamics at a country-scale and its relation with global diversity.

ACKNOWLEDGMENTS: The work of AC Pineda-Peña was supported by the Doctoral Research Training Program "Francisco Jose de Caldas" by the *Departamento Administrativo de Ciencia, Tecnología e Innovación*, COLCIENCIAS, Republic of Colombia and by ERACOL, a scholarship for academic exchange in medicine and the health sciences, between Europe and Latin America. NR Faria is supported by Fundação para a Ciência e Tecnologia under grant agreement no. SFRH/BD/ 64530/2009. The research was supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under the project "Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN)" grant agreement n° 223131; by the Fonds voor Wetenschappelijk Onderzoek – Flanders (FWO) grant G.06.11.09; by the European Union Seventh Framework Programme (FP7/ 2007-2013) under grant agreement no. 278433-PREDEMICS and ERC grant agreement no. 260864.