Phylogeographic Study of HIV-2 Groups A and B Early Epidemics in Western Africa

**BACKGROUND**

- To date ten distinct lineages of HIV-2 have been identified and appear to represent independent zoonotic transfers. These lineages have been termed groups A to D, but only groups A and B have spread in humans (Gei et al. 1994, Damond et al. 2004).
- Group A is mainly found in Western Africa and seems to have originally spread from Guinea-Bissau during the independence war (1963-1976) (Levy 2003).
- HIV-2 group B mainly predominates in Ivory Coast (Panzarini et al. 1999).
- In France, HIV-2 represents 1 to 1.7% of all newly diagnosed HIV infections each year between 2009 and 2012 and more than 1100 HIV-2 infected patients have been included in the French ANRS CO5 HIV-2 cohort since 1994.
- Among patients identified in France, HIV-2 diversity is higher in other Northern European countries with a higher prevalence of HIV-2 group B (35% of all identified viruses) and subtype A2 (25%). This is explained by a different immigration patterns with a strong representation of patients from Ivory-Coast, Mali, Burkina-Faso or Ghana (Damond et al. 2005, Visseau et al. 2016).

**OBJECTIVES**

- We explored the early spread of HIV-2 group A and subtype A2, recently described, in Western Africa using the unique French ANRS HIV-2 cohort database.

**METHODS**

- All HIV-2 pol sequences for which both the time of sampling and patient’s country of birth was known were retrieved from public databases (n=49 and B for groups A and B, respectively) and the ANRS CO5 HIV-2 cohort (n=225 and 68 for groups A and B, respectively).
- All these sequences were sampled between 1986 and 2014.
- All potential recombinant sequences were identified and excluded using RDP 4.0.
- Bayesian MCMC phylogeographic reconstructions were performed for groups A and B under the best fitting combination of evolutionary, demographic and molecular clock models according to Bayes factor analysis using BEAST 1.8.1 (TRACER-G). Bayesian Skyline and a lognormal relaxed molecular clock model.
- The topology of the trees were consistent with maximum likelihood trees obtained using RAxML.
- Because of the large number of sequences sampled in France in the dataset, the patient’s country of birth was used to model the geographical dispersion instead of the sampling country, under the assumption that infections were acquired in the home country of the patients (85% of the patients included in the French ANRS CO5 cohort are born in West Africa).

**RESULTS**

- The Bayesian maximum clade credibility trees are shown in figure 1.
- The distribution of patients according to country of birth and HIV-2 subtype is shown in Table 1.
- The estimated time of the most common ancestors (TMRCA) was 1945 (95% HPD 1935-53) and 1962 (1956-70) for groups A and B, respectively.

**DISCUSSION**

- This phylogeographic study is the first to reconstruct the early subtype A2 and group B dispersal and allows a better understanding of the HIV-2 early epidemics spreading pattern in West Africa.
- The TMRCA observed for HIV-2 group A, 1945 (1935-53), is concordant with previous estimations done by Farlai et al. (1998:1928-45) or Lemey et al. (2014:1924-56) on env and gag genes.
- The TMRCA HIV-2 group B, 1961 (1957-66) overlaps with the previous calculation done by Lemey et al. but is slightly more recent (1965:1955-60).
- Both A2 and B clades firstly emerged in Ivory Coast and diversified later than subtype A1 suggesting that local historical or socio-demographic events may have triggered the dispersal of these viral strains.
- An early founder effect was identified in Senegal and occurred before the Guinea-Bissau independence war. This suggest that HIV-2 group A was already circulating in these two countries before the independence war of Guinea-Bissau that greatly contributed to further dispersal of HIV-2 within and outside West Africa (Levy et al. 2003).
- All these observations need to be confirmed by the env gene analysis within the French ANRS CO5 cohort database. This new dataset will be helpful as more environmental sequences are publicly available contrary to pol gene, especially for Ghana where the group B is highly represented, and as env is well known to exhibit more temporal signal than pol.

**REFERENCES**

- 1. IAME, UMR 1137, Univ Paris Diderot, Sorbonne Paris Cité; INSERM; AP-HP Hôpital Bichat, Virologie, Paris, France; 2. LSHTM, London, United Kingdom, Hôpital Bichat-Claude Bernard, AP-HP, Paris, France; 3. Université Bordeaux, ISPED, Centre INSERM U8997, Epidemiologie Biostatistique, Bordeaux, France; 4. IAME, UMR 1237, Univ Paris Diderot, Sorbonne Paris Cité; INSERM; AP-HP Hôpital Bichat, Maladies Infectieuses et Tropicales, Paris, France.