AIDS Research and Human Retroviruses
© Mary Ann Liebert, Inc.
DOI: 10.1089/AID.2018.0112

HIV-2 infection in a migrant from Gambia: the history of the disease combined with phylogenetic analysis revealed the real source of infection.

Eleonora Cella[1§], Brian T. Foley[2§], Elisabetta Riva[3], Vittoria Scolamacchia[4], Giancarlo Ceccarelli[5,6], Serena Vita[5,6], Marco Iannetta[5], Maria Rosa Ciardi[5], Gabriella D'Ettorre[5], Silvia Angeletti[4]*, Massimo Ciccozzi[1].

[1] Unit of Medical Statistic and Moelcular Epidemiology, University Campus Bio-medico of Rome, Italy.

[2] HIV Databases, T-6, Los Alamos National Laboratory, Los Alamos, New Mexico

[3] Unit of Virology, University Campus Bio-medico of Rome, Italy.

[4] Unit of Clinical Laboratory Science, University Campus Bio-medico of Rome, Italy.

[5] Department of Public Health and Infectious Diseases, Sapienza University of Rome, Policlinico Umberto I, Piazzale Aldo Moro, 00185 Rome, Italy.

[6] Migrant Health Research Organisation (Mi-HeRO) - Centro di Ricerca sulla Salute delle Popolazioni Mobili e Globale, via del Pigneto 3, 00176 Rome, Italy.

[§] These authors equally contributed.

Running Head: Source of HIV-2 infection in a migrant from Gambia

Key words: HIV-2; migrant; origin of infection; phylogenetic analysis

***Corresponding author**

Prof. Silvia Angeletti

Unit of Clinical Laboratory Science

University Campus Bio-medico of Rome

Via Alvaro del Portillo 200

00128 Roma, Italy

++3906225411461

s.angeletti@unicampus.it

Abstract

Human immunodeficiency virus type 2 (HIV-2) infection prevalence is increasing in some European countries. The increasing migratory flow from countries where HIV-2 is endemic has facilitated the spread of the virus into Europe and other regions. We describe a case of HIV-2 infection in a migrant individual in the Asylum Seeker Centre (ASC) in Italy. The patient's virus was sequenced, and found to be a typical HIV-2 genotype A virus. Bayesian evolutionary analysis revealed that the HIV-2 sequence from migrant dated back to 1986 in a subcluster including sequences from Guinea Bissau. This was coherent with the migrant history who lived in Guinea Bissau from his birth until 1998 when he was 13 years old. Monitoring for HIV-2 infection in migrants from western Africa is necessary using adequate molecular tools to improve the diagnosis and understand the real origin of infection.

AIDS Research and Human Retroviruses
HIV-2 infection in a migrant from Gambia: the history of the disease combined with phylogenetic analysis revealed the real source of infection. (DOI: 10.1089/AID.2018.0112)
This paper has been peer-reviewed and accepted for publication, but has yet to undergo copyediting and proof correction. The final published version may differ from this proof.

On November 2015, a 32-year-old Gambian man arrived in Italy and was hosted at the Asylum Seekers Centre of Castelnuovo di Porto, Rome, Italy.

During his residency at the ASC, the patient experienced pneumonia and shingles on the left hemithorax. A panel of blood exams was performed, including HIV tests, which resulted positive only for the p24 band. A further HIV test was scheduled after 8 weeks. Since January 2016, the patient was affected by diarrheal syndrome progressively worsening with the appearance of fatigue, weight loss, anorexia, sweating, arthromyalgia, cough and haemoptysis. He was admitted to the hospital in February 2016. No other disorders were mentioned in his past medical history.

Clinical examination revealed emaciated *facies,* muscular hypotrophy, oral candidiasis associated to hairy leucoplakia and painful abdomen with the deep palpation in the right iliac pit, generalized lymphadenopathy with 1–2 cm sized lymph nodes.

Laboratory investigations revealed pancytopenia with a normocytic anaemia and increased inflammatory markers, as reported in Table 1.

Based on the previous HIV test, the patient was tested a second time for HIV antibodies. HIV-Ag/Ab resulted positive (>12; cut off >1.) (Advia Centaur Systems HIV Ag/Ab Combo assay, Siemens Healthcare Diagnostics) with a detectable HIV-2 RNA viral load (>3000 copies/ml, HIV-2 Real-time RT-PCR, Liferiver). CD4+ T cell count was 98 cells/µL (18.8 %) with a CD4/CD8 ratio of 0.46.

Resistance mutations were evaluated by sequencing the pol gene in the Protease and Reverse Transcriptase regions (Big Dye terminator Cycle sequencing Kit, Applied Biosystem, ABI-3130). Sequence analysis showed HIV-2, subtype A, with susceptibility to all protease inhibitors (PI) and nucleoside reverse transcriptase inhibitors (NRTI) and resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI). These findings were consistent with HIV-2 infection.

A combined antiretroviral therapy (cART) with emtricitabine (200 mg/die) and tenofovir disoproxil (245 mg/die), ritonavir (100 mg/die) boosted darunavir (800 mg/die) was started. Furthermore, trimethoprim-sulfamethoxazole oral administration for *Pneumocystis jiroveci* prophylaxis was prescribed. CD4 cell count improved within

1 month, reaching 148 cells/µL, and 132 cells/µL with a CD4/CD8 ratio of 0.57 after 1 year. Abdominal symptoms disappeared.

At further investigation, patient denied any risk factors for HIV infection, reporting condom use, no surgery or religious practices requiring the use of blood or skin cuts.

Phylogenetic analysis was performed with the aim to evaluate the origin and evolution of the virus and the dynamic of HIV-2 infection in this patient. At this purpose, a 810 bp fragment of the *env* gene was sequenced and analysed.[1]

For phylogenetic analysis, two different dataset were built: the first including the HIV-2 *env* patient sequence plus 29 genotype-specific reference sequences (A1, A2, B, G, H). The second dataset composed by the HIV-2 *env* patient sequence plus 344 genotype A reference sequences. The first dataset was used for typing HIV-2 *env* sequence. The second dataset was to investigate the evolution history of HIV-2 genotype A.

The reference sequences downloaded from the national centre for biotechnology information (https://www.ncbi.nlm.nih.gov/) and from the Los Alamos database (https://www.hiv.lanl.gov/content/index/), were selected based on the following inclusion criteria: 1) sequences already published in peer-reviewed journals; 2) no uncertainty about genotype/subtype assignment; 3) sampling dates were known and clearly established in the original publication.

The sequences were aligned using MAFFT software v.7 (http://mafft.cbrc.jp/alignment/server/), and the alignment was manually refined using Bioedit software.

The phylogenetic signal has been investigated with the likelihood mapping method by analyzing groups of four randomly chosen sequences, called quartets. A quartet has three possible unrooted tree topologies. The three likelihoods are reported as a dot in an equilateral triangle (the likelihood map). A substantial star-like signal (i.e., a star-like outburst of multiple phylogenetic lineages) is indicated by >33% dots falling within the central area, as confirmed by extensive simulation studies. For substitution/saturation analysis assessment, the Xia's test was used with transitions/transversions ratio vs.

divergence graph in DAMBE (http://dambe.bio.uottawa.ca/DAMBE/). The percentage of constant sites and parsimony-info sites were estimated using MEGA7 [2].

The Phylogenetic relationships were analyzed by constructing a Maximum Likelihood phylogenetic tree by MEGA7. The best substitution model (HKY+I+G) was selected by analysis of sequences with the Models tool in MEGA. Tree reliability was assessed by setting bootstrap replicates to 1000. Bootstrap values > 70 were considered significant. The tree was rooted with midpoint rooting and edited using FigTree v1.4.0.

Analysis of the temporal signal and 'clocklikeness' of molecular phylogenies on the second dataset was performed using TempEst. [3] This analysis was performed to evaluate the robustness in terms of molecular clock of the second dataset. The evolutionary rate was estimated on the second dataset by calibrating a molecular clock using known sequences sampling times with the Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v. 1.8.2 (http://beast.bio.ed.ac.uk).[4]

In order to investigate the demographic history, independent MCMC runs were carried out, enforcing both a strict and relaxed clock with an uncorrelated log normal rate distribution and one of the following coalescent priors: constant population size, exponential growth, non-parametric smooth skyride plot Gaussian Markov Random Field (GMRF), and non-parametric Bayesian skyline plot (BSP).[5-7]

Marginal likelihoods estimates for each demographic model were obtained using path sampling and stepping stone analyses. [8-10] Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals, and the best fitting model for each data set was by calculating the Bayes Factors (BF). [8, 11] In practice, any two models can be compared to evaluate the strength of evidence against the null hypothesis (H0), defined as the one with the lower marginal likelihood: 2lnBF < 2 indicates no evidence against H0; 2–6, weak evidence; 6–10: strong evidence, and > 10 very strong evidence. Chains were conducted for at least 100x 10^6 generations, and sampled every 10000 steps for each molecular clock model. Convergence of the MCMC was assessed by calculating the Effective Sample Size (ESS) for each parameter. Only parameter estimates with ESS's of >250 were accepted. The maximum clade credibility (MCC) tree was obtained from the trees posterior distributions, after a 10% burn-in, with the Tree-Annotator software v 1.8.2, included in the Beast package. [4-5] Statistical support for specific monophyletic clades was assessed by calculating the posterior probability (pp>0.90).

The phylogenetic noise investigated by likelihood mapping showed that the percentages of dots falling in the central area of the triangles were 15.7% and 15.7% for the first and second data sets, respectively. Furthermore, since none of the data sets showed more than 33% noise, they had sufficient phylogenetic signal. The percentage of the Parsimony-Informative sites was 44.48% and 68.46% for the first and second dataset respectively; the percentage of the constant sites was 40.46% (first dataset) and 21.16% (second dataset). The phylogenetic signal analysis using a transition/transversion ratio vs. divergence graph and the Xia's test ($p < 0.001$) did not show evidence for substitution saturation (data not shown).

Figure 1 represented the maximum likelihood tree on the first dataset. The isolate from a Gambian migrant clustered with HIV-2 group A reference sequences with a bootstrap value greater than 70%.

Analyzing the temporal signal and 'clocklikeness' of molecular phylogenies on the second dataset a strong correlation between the genetic distance of each sequence to the root of the HIV-2 group A phylogeny and the date of sequence sampling for the second dataset (r=0.614) was found.

In Figure 2, the Bayesian Maximum clade credibility (MCC) tree of HIV-2 group A sequences was showed. The sequences divided in the macro area (continent) are labelled in different colors according to the legend on the left.

Within HIV-2 genotype A, there are several clusters with bootstrap support above 70%. Sequences from Africa intermixed with Europe. The tree suggests a single introduction of the virus into Asia coming from Africa. The red star marks the cluster including the HIV-2 Gambia strain, which is zoomed in the sub-tree to the right of the figure.

This subcluster within the group A viruses dated back to 1986 (HPD 95% 1981 – 2011) and including the Gambia patient plus sequences from Guinea Bissau sampled in 2006.

Comment

HIV-2 infection is mainly endemic in West Africa countries, even if the infections in Europe and in Southwest Asia are increasing. In Africa, the highest prevalence is in Guinea-Bissau (8-10%), a former Portuguese colony, instead the low prevalence is in the neighboring countries, Gambia, Senegal and Guinea. Among them, Gambia has a high rate of infection due to prostitution.

In Europe, Portugal has the highest prevalence of HIV-2 infection, as well as the countries that have been past socioeconomic relationship with this country (i.e. Southwest India). The most common genotype is group A which is the predominant HIV-2 virus in Guinea Bissau and Europe. Due to migration from endemic regions to Italy, there were several infection cases (D'Ettorre *et al.*, 2013).

In the study we investigated the evolutionary history of a HIV-2 infected migrant from Gambia who arrived in Italy in 2016 as an asylum seeker in a CARA.

The strain isolated from Gambian migrant was classified as group A.

The MCC tree on the second dataset highlighted a macroarea distribution except for the region where the migration from endemic region is existing (from Africa to Europe or the other way around). This is probably due to the socioeconomic relationship between former colony country and colonizer country. Another important factor could be to the asylum seek migration.

The clade including the Gambian strain with the Guinea Bissau sequences dated back to 1986 (HPD 95% 1981 – 2011). After a second interview with the migrant, it was possible to find out he was born in Guinea Bissau and lived there with his mother until 1998 when he was 13 years old. Looking for the estimated date of the common ancestor (1986) (HPD 95% 1981 – 2011) it is hypothesized that the Gambian patient may have been infected at birth or during the first years of living because he was born in 1985.

Moreover, HIV-2 infection has been demonstrated to have a slow disease progression [12], this could justify the late onset of the AIDS symptoms and the late timing of the HIV test. Often HIV-2 infected people have low or absent plasma viremia, as well as happened in this case, and the reason for the attenuated virulence of HIV-2 remain unknown [12]. Looking at this case report, we would like to underline, how the combination of classical epidemiological investigation, clinical history of the HIV-2 disease and phylogenetic analysis together can shed light on the most likely route of infection.

Sequence Data.

The sequence of the clinical case has been submitted to GenBank under the accession number MH330317.

References

1.Ciccozzi M, Callegaro A, Lo Presti A, et al. When phylogenetic analysis complements the epidemiological investigation: a case of HIV-2infection, Italy. New Microbiol 2013;36:93-96.

2. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 2016;33:1870-1874.

3. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016 9;2:vew007.

4. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 2005;22:1185–1192.

5. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007;7:214.

6. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 2002;61:1307-1320.

7. Minin VN, Bloomquist EW, Suchard M.A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008;25:1459-1471.

8. Baele G, Lemey P, Vansteelandt S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. BMC Bioinformatics 2013;14:85.

9. Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol 2013;30:239–243.

10. Baele G, Lemey P. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. Bioinformatics 2008;29: 1970–1979.

11. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995;90:773-795.

12. MacNeil A, Sankale JL, Meloni ST, Sarr AD, Mboup S, Kanki P. Long-term intrapatient viral evolution during HIV-2 infection. J Infect Dis 2007;195:726-733.

Figure legends

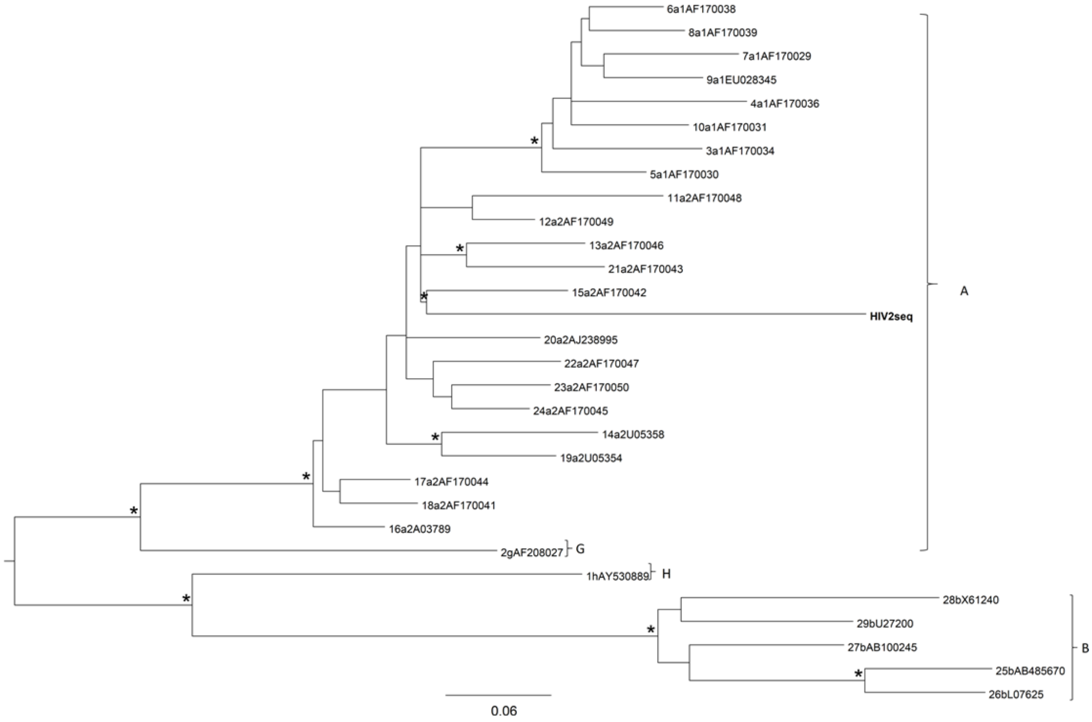

Figure 1. Phylogenetic relationships of the HIV-2 sequence isolated from a migrant from Gambia (in bold) with the subtype-specific reference sequences downloaded from NCBI sequence database (https://www.ncbi.nlm.nih.gov/). The reference sequences used in the analysis are showed in the tree with their original accession numbers. The asterisks (*) along a branch represent significant statistical support for the clade subtending that branch (bootstrap support >70%). The scale bar indicates 0.06 nucleotide sequence divergence.

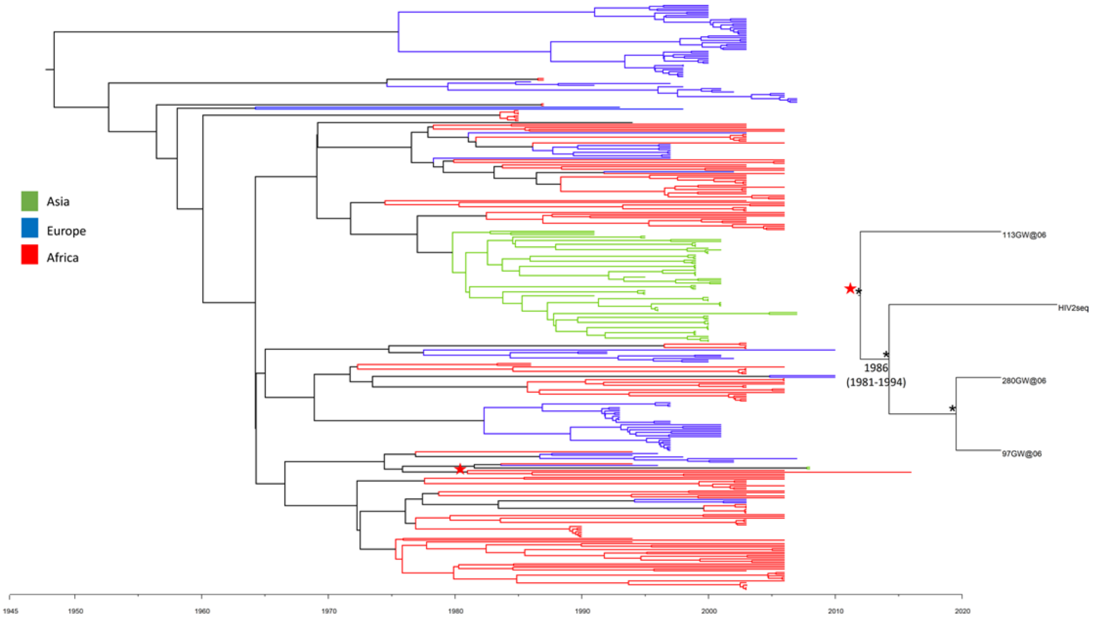

Figure 2. Bayesian Maximum clade credibility (MCC) tree of HIV-2 group A. The sequences divided in the macro area (continent) are labelled in different colors according the legend on the left. The asterisks (*) along a branch represent significant statistical support for the clade subtending that branch (bootstrap support >70%). The red star marks the cluster including the HIV-2 Gambia strain zoomed in the sub-tree to the right side.