HIV-1 subtype C in Ethiopia: genotypic and phenotypic variation
Abebe, A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER VI

Timing of the introduction into Ethiopia of the subcluster C' of the HIV-1 subtype C

Almaz Abebe, Vladimir V. Lukashov, Tobias F. Rinke de Wit, Bitew Fisseha, Belete Tegbaru, Aletta Kliphuis, Girma Tesfaye, Hailu Negassa, Arnaud L. Fontanet, Jaap Goudsmit, and Georgios Pollakis

Submitted for publication
ABSTRACT: To analyze genetic diversification within the HIV-1 C' population in the course of the epidemic in Ethiopia, we analyzed 165 env gp120 V3 sequences obtained between 1988 and 1998. We observed a highly significant positive correlation between sampling years of individual sequences and their synonymous distances to the reconstructed common ancestor of the Ethiopian C' epidemic. The extrapolation of the regression line of synonymous distances back to the date when no synonymous heterogeneity was present among the Ethiopian HIV-1 C' population allowed to estimate 1982 (95% CI: 1980-1983) as the year of the onset of HIV-1 C' genetic diversification and expansion in Ethiopia. These results are in agreement with retrospective epidemiological and serological data, which demonstrated the absence of HIV-1 epidemic in the Ethiopian population in the early 1980's.

RESULTS AND DISCUSSION: More than 15 years have elapsed since the HIV/AIDS epidemic in sub-Saharan Africa became evident. While all known HIV-1 genetic subtypes are found in Africa, subtype C is becoming increasingly prominent in terms of its geographical distribution and the total number of infections caused, especially in the eastern and southern parts of the continent [1-11].

In Ethiopia, retrospective serological studies revealed the absence of HIV-1 in rural or urban populations prior to 1984 [12,13]. The first case of HIV-1 infection has been registered in 1984 in the capital city Addis Ababa, when serum samples of 167 hospitalized patients have been tested for anti-HIV-1 antibodies [13]. The first AIDS cases in Ethiopia have been registered in 1986 in hospital in Addis Ababa [14]. Since 1986, an increasing prevalence of HIV-1 has been demonstrated. In 1994, data from eleven urban blood banks showed that the HIV-1 prevalence varied from 5% to 20%, being 6.6% in Addis Ababa (National Blood Transfusion Service, Ethiopian Red Cross Society, 1994, unpublished). Among pregnant women in Addis Ababa, 11-13% was found to be HIV-1 seropositive in 1991 [15], compared to 18% in 1996 [16]. Among commercial sex workers (CSW), HIV-1 prevalence increased from 6% in 1985-'86 [17] to 74% in 1998 (Aklilu M. et al., submitted).

Previous molecular epidemiological studies revealed that the HIV-1 epidemic in Ethiopia is caused by HIV-1 subtype C viruses [18-28]. Phylogenetic analysis of Ethiopian env gp120 V3 sequences demonstrated circulation of two genetically different subtype C virus populations in Ethiopia, designated C and C' [18]. It has been shown, that both C and C' viruses have been circulating over the last decade among the same risk groups and geographical areas [Abebe A, et al. submitted].
In the previous study, we obtained and genetically characterized the earliest HIV-I strains that have been present in Ethiopia in 1984 and 1985 [Abebe A. et al. submitted].

Subsequently, we analyzed the evolution of the Ethiopian HIV-1 C population over a period of 1984-1998, and observed a significant increase of synonymous distances of HIV-1 C sequences to their reconstructed common node in the course of the epidemic. Our analysis of the relation between synonymous distances of HIV-1 C sequences to the common node and their sampling years, supported epidemiological and serological data that current HIV-1 C epidemic in Ethiopia is resulting from virus introduction and subsequent expansion in the Ethiopian population in the early 1980's [13. Abebe A. et al. submitted].

In the present study, we performed similar analysis for the C' virus population in Ethiopia. A set of 165 C' V3 sequences was obtained from serum samples collected in Ethiopia over a period between 1988 and 1998 under the framework of the national-wide HIV-1 surveillance project. The samples were obtained from CSW, blood donors, hospital patients, pregnant women, and factory workers in several Ethiopian cities and towns: Addis Ababa, Arba-Minch, Assab, Dire Dawa, Dessie, Gondar, Jimma, Akaki, and Wonji.

To analyze whether the consensus sequence of the V3 region of the subtype C' epidemic in Ethiopia is changing in the course of the epidemic, we calculated consensus sequences for seven periods of the sequences obtained in this and earlier studies [25]; consensus 1988 (n = 23), 1989-90 (n = 11), 1992 (n = 17), 1995 (n = 39), 1996 (n = 22), 1997 (n = 56) and 1998 (n= 15) (Figure 1). For each sequence position, the most often amino acid was included in the consensus, even if it was not present in the majority of sequences. Differences between the consensus sequences of the seven periods of the epidemic were observed at eleven positions, and all but four different amino acids were seen during two or more consecutive time-periods. At the sequence positions 12, 21, 66 and 70 an amino acid that was present among early samples, was changed among late samples. The valine (V) at position 12 of the

| Consens 88-98: IVIRSENLTNNAKIIIIVHLKEPVEIVCTRPTNTRKSMRIGPGQTYTFATGDIDIRQAHCNISEKTWNETLQEVGKDLQEMFPNKTI |
| Consens 88: V--K--- ---Y- |
| Consens 89-90: K--D-V--- ---K- |
| Consens 92: ---V--- ---K--- |
| Consens 95: ---Q--- ---K--- |
| Consens 96: ---Q--- ---K--- |
| Consens 97: ---E--- ---K--- |
| Consens 98: ---Q--- ---K--- |

Figure 1: Predicted consensus amino acid sequences of the envelope gp120 V3 region for seven periods of the subtype C' epidemic in Ethiopia. The sequences are shown in comparison with the overall C' Ethiopian consensus sequence. Dashes indicate amino acid identities.
early period, 88 to 92, changed to an Alanine (A) for the period of 95 to 98. Likewise the 88-95 Lysine (position 66) changed to a Glutamic acid (E) in 96–97. The Lysine (K) (position 21) of the period 88 to 90 changed to a Glutamic acid in 92 to 95 turned back to a Lysine in 96 to finally become Glutamic acid for 97-98 and the glutamic acid of 88 – 90 became a Lysine during 92–96 only to turn back during 97-98.

All sequences were aligned, and their most recent common ancestor of the Ethiopian C' virus population was reconstructed as the common node of the C' cluster in the neighbor-joining phylogenetic tree, constructed by using PHILIP package. Ethiopian C sequences and sequences of HIV-1 subtype other than C were used as an outgroup. The synonymous distances of the 165 sequences to their common node were calculated ($D_s$, Jukes-Cantor correction). Additionally, the mean synonymous sequence distances per each sampling year ($\pm$ SE) were calculated. All statistical calculations were done by using the SPSS/PC+ software (version 5.0, SPSS Inc., Chicago, Illinois, USA). The relationship (correlation) between sampling years of sequences and their synonymous distances to the most recent common ancestor was examined by using linear regression analysis [29,30]. Each sequence was considered to be statistically independent. The distance data were studied by using linear regression analysis in two ways. In the first method, the distances of all individual sequences to the common node were analyzed in relation to their sampling year (all sequences equally contributed to the analysis). In the second approach, the mean distances calculated per year were used (all years equally contributed to the analysis). Ninety-five percent confidence intervals for the time of introduction of C' viruses into Ethiopia has been calculated (CI for X-intercept).

Regression analysis of evolutionary distances of 165 Ethiopian C' V3 sequences revealed that there is a highly significant ($p<0.00000001$) positive correlation between sampling years of individual sequences and their synonymous distances to the reconstructed common ancestor of the Ethiopian C' epidemic (Figure 2). The extrapolation of the regression line of synonymous distances back to the date when no synonymous heterogeneity was present among the Ethiopian HIV-1 C' population (X-intercept) allowed to estimate 1982 (95% CI: 1980-1983) as the year of the onset of HIV-1 C' genetic diversification in the Ethiopian population. Similar results were obtained when regression analysis was based on the mean synonymous distances of all sequences sampled in each year (Figure 1). This approach resulted in a marked increase of correlation ($r=0.98, r^2=0.97$) between the mean synonymous distances of sequences and their sampling years. The synonymous evolution rate of C' viruses in
our study was similar to those observed earlier for HIV-1 subtype B populations circulating in the US and among injecting drug users and homosexual men in the Netherlands (0.004 versus 0.004; 0.005; 0.005) respectively [29,30].

**Figure 2:** Synonymous distances of 165 Ethiopian C' V3 sequences to the reconstructed common ancestor of the Ethiopian C' epidemic in relation to sequence sampling years. The regression analysis is performed for all 165 individual sequences (closed circles, solid regression line) as well as for the mean distances per sampling year (open circles, dotted regression line, error bars are shown). Dashed lines indicate 95% confidence intervals calculated based on 165 individual sequences. Statistics: for individual sequences - r=0.45, r²=0.20, slope 0.004, X-intercept - 1982, p<0.00000001; for the mean distances - r=0.98, r²=0.97, slope 0.007, X-intercept - 1982, p<0.00001.

These results, together with our earlier study on HIV-1 C epidemic in Ethiopia [Abebe A, et al. submitted] indicate that, the trends in genetic heterogeneity among the C and C' virus populations are supporting the concept that the current HIV-1 epidemic in Ethiopia is resulting from at least two HIV-1 introductions and subsequent virus expansion in the early 1980's. Our genetic results are in agreement with retrospective epidemiological and serological data on the absence of the HIV-1 epidemic in Ethiopia before 1984-1985.
REFERENCES


ACKNOWLEDGEMENTS

This study is part of the Ethio-Netherlands AIDS Research Programme (ENARP), a collaborative effort of the Ethiopian Health and Nutrition Research Institute (EHNRI), the Amsterdam Municipal Health Service (GG/GD), the Central Laboratory of the Netherlands Red Cross Blood Transfusion Service (CLB) and the Academic Medical Center of the University of Amsterdam (AMC). ENARP is financially supported by the Netherlands Ministry of Foreign Affairs and the Ethiopian Ministry of Health (MOH) as a bilateral project. The authors would like to thank Mr. Tesfaye Mebratu for his voluntary technical assistance and Bloodbank technicians of the various towns for collecting and sending the samples included in the study.