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 IV

Identification of a genetic sub-cluster of HIV-1 subtype C (C’) widespread in Ethiopia

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

AIDS Res Hum Retrovir 2000; in press
Abstract: We and others have previously shown that subtype C is the predominant HIV-1 subtype and the major cause of AIDS in Ethiopia. The present study shows that subtype C in Ethiopia has a genetic sub-cluster, designated C' that does not increase in frequency, nor spreads geographically over the period 1988 (%C' = 23/33) to 1996/7 (%C' = 26/50). There is no association of the HIV-1 subtype C or sub-cluster C' with geographic location, time of sample collection or risk group in Ethiopia. Out of 105 randomly collected samples representing 7 different towns in Ethiopia, all but two (one subtype A from Addis Ababa, 1997 and one subtype D from Dessie, 1996) belong to subtype C.

Results and discussion: Human immunodeficiency virus type one (HIV-1) subtypes are distributed unevenly across Africa nations. In East and Central African countries such as Uganda, Kenya and Tanzania, the HIV-1 epidemic involves mainly two HIV-1 subtypes, A and D. In contrast, subtype C has dominated the rapidly expanding epidemic in Botswana, South Africa and Ethiopia. HIV-1 subtype C is on the rise, possibly gradually replacing subtype D viruses in Eastern African countries.

The relative roles played by virological, behavioral, and host determinants in the epidemic expansion of any particular HIV-1 subtype are unclear. Careful surveillance of genetic subtypes in a given population is presently a particularly important approach for better understanding the biological properties of different HIV-1 subtypes.

The first Ethiopian HIV-1 positive sera were detected in 1984 and the first Ethiopian AIDS case was reported in 1986 in Addis Ababa the capital city with presently more than 3 million inhabitants. A national surveillance performed in 1988 amongst commercial sex workers in 23 towns and cities in Ethiopia revealed an HIV-1 prevalence of 1% to 38%. In 1994 the sero-prevalence was 7% among blood donors. Sentinel surveillances performed in 1995, 1996 and 1997 in Addis Ababa report HIV-1 prevalence ranging from 14 to 20% in pregnant women. Recent studies performed in 1997 and 1998 reveal an HIV-1 prevalence of 45-74% in commercial sex workers in Addis Ababa (Aklilu M., personal communication). Finally, sequence data on sera and plasma samples collected in Addis Ababa, demonstrated that the majority of the circulating viruses belong to subtype C. These Ethiopian subtype C sequences differ slightly from the consensus C sequence and there was some hint of the presence of a separate sub-cluster within the main C group. Nevertheless, the presence of such sub-cluster was not supported by a significant bootstrap value in phylogenetic tree analysis. To assess the geographical distribution in more detail and the possible influx of
HIV-1 subtype(s) in the whole of Ethiopia, 105 serum samples were analysed which have been collected from seven different towns and two risk groups. Figure 1 shows the map of Ethiopia, the number of serum samples sequenced, the town and the year of sample collection.

![Map of Ethiopia and Eritrea](image)

<table>
<thead>
<tr>
<th>Town</th>
<th>Number of samples collected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1988(C')</td>
</tr>
<tr>
<td>Addis Ababa</td>
<td>7(4)</td>
</tr>
<tr>
<td>Gondar</td>
<td>7(5)</td>
</tr>
<tr>
<td>Dessie</td>
<td>7(1)</td>
</tr>
<tr>
<td>Dire Dawa</td>
<td>9(2)</td>
</tr>
<tr>
<td>Jimma</td>
<td>7(2)</td>
</tr>
<tr>
<td>Arba Minch</td>
<td>9(7)</td>
</tr>
<tr>
<td>Assab</td>
<td>7(2)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>53(23)</strong></td>
</tr>
</tbody>
</table>

**Figure 1:** Map of Ethiopia and Eritrea. The numbers in the map indicate the different towns included in the study. The samples shown in the table were collected from commercial sex workers in 1988 and blood donors in 1996/7. The numbers in the parenthesis indicate the number of isolates that belong to the sub-cluster C'.
The procedures of HIV-1 RNA isolation, reverse transcription and direct sequencing were described earlier. The C2-V3 region of the envelope gp120 glycoprotein gene was sequenced and sequence alignment was performed manually according to the Los Alamos reference-sequences used for subtyping. The nucleotide alignments were subjected to phylogenetic tree analysis using neighbor joining and maximum likelihood method implemented in the Phylip package programs and using kimura-2-parameter distances. The bootstrap option in the MEGA program was used to determine the reliability of the clusters in the phylogenetic trees. A bootstrap value equal or greater than 75% was considered significant. The synonymous and non-synonymous nucleotide substitution distance matrix was generated using the MEGA program by the Jukes and Cantor method. Multivariate principal co-ordinate analysis was done with the PCOORD software and amino-acid sequence comparisons were done using the VESPA program.

**Figure 2:** Phylogenetic tree analysis of Ethiopian HIV-1 sequences by neighbor-joining algorithm (17). The sub-cluster is indicated as sub-cluster C' in the tree. The HIV-1 subtype A (AA97206) and subtype D (DE96050) isolates found in this study are shown, clustering with the subtype A (TP95001) and the subtype D (KS39671) from our previous studies (19,21). Sequences have been indicated by codes: AA = Addis Ababa, AM = Arba Minch, AS = Assab, DD = Dire Dawa, DE = Dessie, JM = Jima and GO = Gondar. The first two digit number following these codes indicate the year of sample collection, the next three digits indicate sample number. The GenBank accession numbers are AF245518 to AF2456613. Numbers by the branches represent bootstrap values out of a 100 replications.
The phylogenetic analysis depicted in Figure 2 clearly shows that subtype C is widely distributed and dominates the HIV-1 epidemic in Ethiopia. Among the 105 samples, two sequences clustered with previously published subtype A and subtype D sequences from Addis Ababa. Sample AA97026 of Addis Ababa showed homology with TP95001 (subtype A, pregnant women, Addis Ababa, 1995), while DE96050 from Dessie clustered with KS39671 (subtype D, commercial sex worker, Addis Ababa, 1997). All serum samples collected in 1988 contained subtype C virus. Although both HIV-1 subtype A and D isolates were collected in 1996/7, there is evidence for an earlier introduction of HIV-1 subtype A in Ethiopia, in or even before 1991.

Figure 3: Result of the multivariate principal co-ordinate analysis (PCOORD) for the main group C and the sub-cluster C'. Axes are the two dimensions that were first extracted; together they cover 24% of the total difference between the two groups. The first ten axes cover 53% of variation. The arrow points the subtype C reference C2220. Two reference sequences of subtype A and D were included, noted as such.
Within the Ethiopian sequences a sub-cluster (named C') could be identified with a significant bootstrap value (bootstrap 77%). The Ethiopian isolate C2220 from 1986, used as a C reference, does not belong to the C' sub-cluster. The prevalence of both the sub-cluster C' and the main group C viruses appears equally distributed, as this study indicates that 48% of the analysed serum samples contain a C' virus and 52% a C virus. The sub-cluster was confirmed by maximum likelihood phylogenetic analysis (data not shown) and by multivariant principal co-ordinate analysis (Figure 3).

The Ethiopian sub-cluster C' viruses are grouped together, independent of any variable considered in this study, which includes geography, risk group and time of sample collection. The two groups co-circulate with similar prevalence, although the rate of synonymous and non-synonymous nucleotide variation amongst the sequences of the C group is higher than in the C' group. These differences are statistically significant (p<0.0001) for the two time points of sample collection.

The VESPA software was used for the amino acid sequence comparison of the Los Alamos data base (3) subtype C sequences (LsAC) and the two subgroups C (Eth C) and C' (Eth C') circulating in Ethiopia. Both groups C and C' show significant amino-acid differences when compared to the database or to each other but the main group C is genetically closer to the data base sequences than to the sub-cluster C' ones (Figure 4).

![Figure 4: VESPA supported amino-acid sequence comparison between the Los Alamos data-base subtype C (LsAC) sequences and the C (I) and C' (II) groups in Ethiopia indicated as Eth C' and Eth C' respectively.](image)

The presence in the C' group of a Lysine (K) at position 304 instead of a Glutamic acid (E) affects the number of positive charges in V3 loop and a Valine (V) instead of an Asparagine (N) at position 294 leads to the loss of a potential N-glycosylation site. We do not have evidence that there are significant biological differences between the groups C and C', although it is known that the number of charges and glycosylation in the V3 loop can affect...
cellular tropism and neutralisation ability of antibodies. Further studies and follow-up of the epidemic are needed to answer these questions.

We conclude that viruses circulating in Ethiopia over the last 10 years cluster with the main subtype C, but a significant sub-cluster C’ was noted in multiple analyses. This sub-cluster of subtype C (C’) was in a fifty-fifty equilibrium with the main subtype C in Ethiopia in the last 10 years of the epidemic.
References


18. Kumar, S., Tamura, K., and Nei, M. 1993. Molecular Evolutionary Genetics Analysis (MEGA), version 1.01, Institute of Molecular Evolutionary Genetics, Pennsylvania State University, USA.


Acknowledgements

This study is part of the Ethio-Netherlands AIDS Research Project (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.