Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification

Almaz Abebe, Vladimir V. Lukashov[a], Georgios Pollakis[a], Aletta Kliphuis[a], Arnaud L. Fontanet[b], Jaap Goudsmit[a] and Tobias F. Rinke de Wit

Objective: To trace the introduction of HIV-1 subtype C into Ethiopia based on virus diversification during the epidemic.

Design: A set of 474 serum samples obtained in Ethiopia in 1982–1985 was tested for HIV-1. HIV-1 *env* gp120 V3 and *gag* or *pol* regions were sequenced and analysed together with sequences from later stages of the epidemic.

Results: None of 98 samples from 1982–1983, one of 193 samples from 1984, and one of 183 samples from 1985 were HIV-1 positive. Phylogenetic analysis of virus sequences from positive samples revealed that they belong to the Ethiopian C, and not the C′, cluster. Analysis of 81 Ethiopian C V3 sequences from 1984–1997 revealed that the consensus sequence of the Ethiopian epidemic has been stable over time. Both the 1984 and 1985 V3 sequences, in contrast with three out of 27 (11%) of the 1988 and none out of 51 of the 1992–1997 sequences, had no synonymous substitutions compared to the reconstructed common ancestor of the Ethiopian C viruses. A highly significant correlation between sampling years of the V3 sequences and their synonymous distances to the common ancestor was demonstrated.

Conclusions: The increasing genetic heterogeneity together with stable consensus sequence of the Ethiopian HIV-1 C population demonstrates that evolution of the virus population is characterized by an unbiased expansion around a stationary consensus. Based on the rate of synonymous diversification of HIV-1 strains within the Ethiopian population, we were able to estimate 1983 (95% confidence interval, 1980–1984) as the year of HIV-1 C introduction into Ethiopia. © 2001 Lippincott Williams & Wilkins

AIDS 2001, **15**:1555–1561

Keywords: Ethiopia, HIV-1 epidemic, HIV-1 genetic characterization, molecular clock, subtype C

Introduction

Like most countries in Sub-Saharan Africa, Ethiopia has experienced a severe HIV-1/AIDS epidemic during the last 15 years. Intensive retrospective seroepidemiological studies, the results of which are summarized in Table 1, have revealed the absence of HIV-1 in rural or urban Ethiopian populations prior to 1984 [1,2].

From the Ethiopian–Netherlands AIDS Research Project (ENARP) at the Ethiopian Health and Nutrition Research Institute (EHNRI), Addis Ababa, Ethiopia, the [a]Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands, and the [b]Division of public Health and Environment, Municipal Health Service, Amsterdam, the Netherlands.

Note: Almaz Abebe and Vladimir V. Lukashov contributed equally to this work.

Requests for reprints to: A. Abebe, Ethiopian Health and Nutrition Research Institute, PO box 1242, Addis Ababa, Ethiopia.

Received: 23 May 2000; revised: 6 November 2000; accepted: 24 April 2001.

Table 1. Early HIV-1 seroprevalence in Ethiopia (1982–1987): summary of data published earlier [2] and those from the present study.

Year	Risk group	Location	Serum samples tested (n)	HIV+ [n (%)]
1982	Hepatitis B	Addis Ababa	44	0 (0%)
1982–1983 [2]	Hospital patients	Addis Ababa, Asmara, Harar, Yirgalem, Nekemte	500	0 (0%)
1983	Hepatitis B	Addis Ababa	54	0 (0%)
1983 [2]	Blood donors	Addis Ababa	459	0 (0%)
1984	Hepatitis B	Addis Ababa	193	1 (0.5%)
1984 [2]	Bell's palsy + controls	Ethiopia	267	2 (0.7%)
1985	Hepatitis B	Addis Ababa	183	1 (0.5%)
1985–1987 [2]	Liver diseases	Addis Ababa	528	13 (2.5%)

The first case of HIV-1 infection in Ethiopia was registered in 1984 in the capital city Addis Ababa, when serum samples of 167 hospitalized patients were tested for anti-HIV-1 antibodies [2]. The first AIDS cases in Ethiopia were registered in 1986 in hospitals in Addis Ababa [3]. In 1986–1988, a high prevalence of HIV-1 was demonstrated along the main trading roads of the country, estimated at 13% among truck drivers (n = 468) [4], and 17% among commercial sex workers (CSW) (n = 6234; 23 urban areas) in 1988 [5]. In 1994, data from 11 urban blood banks showed that the prevalence of HIV varied from 5% to 20%, being 6.6% in Addis Ababa (National Blood Transfusion Service, Ethiopian Red Cross Society 1994, unpublished data). Between 11 and 13% of antenatal care attenders in Addis Ababa were found to be HIV-1 seropositive in 1991 [6] compared with 18% in 1996 [7]. A recent study among CSW in Addis Ababa demonstrated a HIV prevalence of 74% [8].

We and others have reported that the HIV-1 epidemic in Ethiopia is caused by subtype C viruses [9–19] – up to 99% of all HIV-1 strains belong to this subtype [9,17,20]. Among the Ethiopian HIV-1 subtype C strains, two virus clusters can be distinguished based on the phylogenetic analysis of their *env* genes: these are designated as main C group and subcluster C′ [9]. It has been shown that both main C group and subcluster C′ viruses have been circulating during the last decade among the same risk groups and in the same geographical areas [20].

In this study, we report the identification of two HIV-1 positive serum samples from Addis Ababa obtained from HIV-1 seropositive individuals sampled in 1984 and 1985. To determine whether these two viruses are genetically close to the founder virus introduced originally into Ethiopia, we obtained and analysed genetic information for these two earliest Ethiopian viruses and compared it with sequence data obtained in later years from the same geographic area.

Materials and methods

Clinical samples

A total of 474 serum samples, obtained during 1982–1985 from patients who visited the former National Research Institute of Health Virology laboratory for routine diagnosis of hepatitis B virus (HBV) infection, were used for our study. These 474 samples were tested for the presence of HIV-1 antibodies by ELISA (Vironostika Uniform II, Organon Teknika, Boxtel, The Netherlands) and Western Blot (HIV Blot version 2.2, Genelabs Diagnostic Biotechnology, Redwood City, California, USA) according to the manufacturers' instructions.

Samples selected for the linear regression analysis (n = 81) belonged to the main HIV-1 C group [9,20] and were obtained in the period 1984–1997. Sera were collected from: HBV/HIV-1 positive patients (n = 2), Addis Ababa 1984 and 1985; an AIDS patient (n = 1), Addis Ababa 1986; CSW, pregnant women and blood donors (n = 26), Addis Ababa 1992–1995; and CSW and blood donors of seven different Ethiopian towns including Addis Ababa (n = 52), 1988 and 1996/1997.

Sequencing

For HIV-1 positive samples, reverse transcription (RT)–PCR of three genomic regions (*env* gp120 V3, *gag* p17/p24, and *pol* protease/partial *RT*) was performed. Briefly, RNA was extracted from 100 μl serum by the silica-based method [21]. After washing and elution from the silica with 100 μl sterile water, 10 μl of the eluate was used in a RT reaction with avian myeloblastosis virus reverse transcriptase. A 287 bp region of the HIV-1 *env* V3 region was amplified with C2-V3 primers and further amplified by nested PCR. The outer primers were 3′-V3-NOT (5′-GCGCGGCCGCCCCCTCTACAATTAAAACTGTG-3′) and 5′-V3-NOT (5′-GCGCGGCCGCACAGTACAATGTACACATGG-3'); the inner primers were 5′-V3-SP6 (5′-GATTTAGGTGACATATAG-

3′) and 3′-V3-T7 (5′-TAATACGACTCACTAT AGGG-3′), as described earlier [22–24]. For amplification of a 743 bp fragment of the *gag* gene (complete p17 and partial p24), outer primers 3′-SK39 (5′-GCATTCTGGACATAAGACAAGGACCAAA-3′) and 5′-GAG-1 (5′-GCGAGAGCGTCAGTATTAAG C-3′) and inner primers 5′-GAG-2-SP6 (5′-GG GAAAAATTCGGTTAAAGGCC-3′) and 3′-GAG-AE3-T7 (5′-TAGGACCCTAATTTATTTTATCA-3′), were used as described earlier [25,26]. For the amplification of the *pol* region (1178 bp including the protease and 297 amino acids of *RT*), the first PCR reaction was performed with the outer primers 5′-protFM (5′-CAAGGGAAGGCCAGGGAATTT-3′, HxB2 positions 2111–2130) and 3′-halfRT (5′-TGACCCATCAAAAGACTTAATAGCAGAAATA -3, HxB2 positions 3505–3535). Subsequently, DNA was amplified in three nested PCR (fragments A, B and C), covering the amplified region. The inner primers used were: for fragment A, 5′-SP6-prot/RT (5′-CTTTAACTTCCCTCAGATCACT-3′, HxB2 positions 2242-2263) and 3′-T7prot (5′-CCTATT GAAACTGTACCAGTA-3′, HxB2 positions 2558-2578); for fragment B, 5′-SP6p66/out (5′-GACCTA CACCTGTCAACATAAT-3′, HxB2 positions 2484–2505) and 3′-endprotT7 (5′-TGGAAAGGATCAC CAGCAATATT-3′, HxB2 positions 3005–3027); for fragment C, 5′-SP6p66 (5′-AGATATCAGTACA ATGTGTT-3′, HxB2 positions 2975-2994) and 3′-halfpol (5′-AAGCAGAGCTAGAACTGGCAGA-3′, HxB2 positions 3441–3462) using the conditions described earlier [27].

Both strands of the nested-PCR fragments were sequenced directly by using the SP6 and T7 primers. Sequencing was performed with Taq dye primers (Applied Biosystems, Foster City, California, USA) and the ThermoSequenase fluorescence-labelled primer cycle-sequencing kit (Amersham International, Little Chalfont, UK). The sequencing reaction products were analysed on an automatic DNA sequencer (model 373A Applied Biosystems).

Sequences obtained in this study have been submitted to GenBank as: hepatitis patients' strains, AF307298, AF30799; CSW's strains, AF245518–AF245521, AF245524, AF245535, AF245538, AF245544, AF245551–AF245554, AF245557–AF245560, AF245562-AF245565, AF245572-AF245578, AF245583, AF245595, AF245597–AF245600, AF245609, AF245610, U88812, U88753, U88758, U88767, U88788, U88790, U88791, U88793, U88796; blood donors' strains, AF245525, AF245527, AF245528, AF245530–AF245534, AF245537, AF245545, AF245549, AF245550, AF245556, AF245568, AF245570, AF245571, AF245579–AF245582, AF245588–AF245590, AF245592–AF245594, AF245602, AF245607, AF245611, AF245612, U88735, U88764, U88774, U88785, U88786, U88815, U88819-U88821; pregnant women's strains, U88750, U88757.

Sequence analysis

Nucleotide sequences of the early samples obtained in this study were aligned manually together with 79 main C virus group sequences obtained from Ethiopia in earlier studies (1986 and 1992–1995 sequences were from Addis Ababa; 1988, 1996/1997 sequences were from other urban areas including Addis Ababa [9,15,20]). Positions containing an alignment gap were excluded from pair-wise sequence comparisons. The most recent common ancestor, or the founder virus, for the Ethiopian subtype C, main C virus epidemic was reconstructed as the common node of 81 Ethiopian main C virus sequences using several phylogenetic methods. Phylogenetic trees were constructed by using the neighbour-joining and maximum-likelihood algorithms as implemented in PHYLIP package (NEIGHBOR and DNAML, respectively) (http://evolution.genetics.washington.edu/phylip.html). The DNAML method was based on empirically found base frequencies and transition/transversion ratios, considering different rates of evolution at different positions. The neighbour-joining method was based on gamma distances for the Jukes–Cantor model which considers different evolution rates at different nucleotide sites. For both methods, reference sequences of HIV-1 subtypes other than C, provided by the Los Alamos database (http://hiv-web.lanl.gov) were used to root the trees. Subtype C sequences from other countries and Ethiopian main C group and subcluster C′ sequences obtained earlier were also included. Subsequently, neighbour-joining phylogenetic trees based on synonymous evolutionary distances (D_s, Nei–Gojobori method with Jukes–Cantor correction) were built by using the MEGA package [28]. For all methods, evolutionary distances of individual sequences to the most recent common ancestor, which was calculated as the common node of all Ethiopian main C viruses, were calculated and analysed by using several statistical approaches.

Statistical methods

All statistical calculations were performed by using the SPSS/PC+ software (version 5.0, SPSS Inc., Chicago, Illinois, USA). The relationship (correlation) between sampling years of individual HIV-1 main C sequences and their D_s to the most recent common ancestor was examined by using linear regression analysis. Each sequence was considered to be statistically independent and the errors of sampling time axis were set to zero. 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 analysed in relation to their sampling year (all sequences contributed equally to the analysis). In the second approach, the mean distances calculated per

year were used (all years contributed equally to the analysis). For each analysis, 95% confidence intervals (CI) for the time of virus introduction into Ethiopia was calculated (CI for X-intercept).

Results

A total of 474 serum samples, collected between 1982 and 1985 from Ethiopian individuals, were tested by ELISA and Western blot for HIV-1 antibodies. None of the samples obtained in 1982 ($n = 44$) or 1983 ($n = 54$) were found to be HIV-1 positive. One out of 193 (0.5%) samples from 1984, and one out of 183 (0.5%) from 1985 were found to contain anti-HIV-1 antibodies (Table 1). These samples were obtained from males aged 26 and 41 years, respectively.

From these two samples, virus-specific RNA was isolated and RT–PCR of the HIV-1 *env*, *pol*, and *gag* genes was performed. Amplification of the gp120 V3 region was possible for both samples, but the *gag* region was amplified successfully for the 1984 sample only, and the *pol* region was amplified for the 1985 sample only. As the *gag* and *pol* regions are, in general, more highly conserved than the *env* V3 region, it is likely that the negative amplification results are related to the longer size of these fragments. The predicted amino acid sequences of the V3 regions of the 1984 and 1985 samples are shown in Fig. 1. Phylogenetic analysis revealed that both the 1984 and 1985 viruses belong to the main C group of Ethiopian subtype C viruses in the gp120 V3 region (Fig. 2). The *gag* region of the 1984 and the *pol* region of the 1985 samples also clustered with the main C sequences and not with the subcluster C′ sequences (data not shown).

To analyse whether the consensus sequence of the V3 region of subtype C viruses in Ethiopia is changing during the course of the epidemic, we calculated consensus sequences for six periods of the epidemic based on 81 sequences obtained in this and earlier studies [9,15,20]: consensus 1984–1986 (based on three sequences sampled between 1984 and 1986); 1988 (27 sequences sampled in 1988); 1992 (six sequences); 1995 (20 sequences); 1996 (19 sequences); and 1997 (six sequences) (Fig. 1). For each sequence position, the most commonly occurring amino acid was included in the consensus, even if it was not present in the majority of sequences (question marks represent cases when two or more amino acids were the most common at this position). Our analysis revealed that during the HIV-1 epidemic in Ethiopia, the consensus V3 sequence has remained stationary. We observed that although there were differences among the consensus sequences present during different time periods of the epidemic, at no position did an amino acid that was absent or in minority among early samples become consistently dominant among later samples. Differences between the consensus sequences of the six periods of the epidemic were observed at 14 positions, and all but one of the different amino acids were seen in one consensus sequence only. At only one position (position 36) was an amino acid difference seen in the consensus sequences of two periods of the epidemic. At this position, the early consensus sequences (1984–1986 and 1988) had E (glutamic acid), whereas the consensus sequences of 1992 and 1995 had K (lysine) as the prevalent amino acid. However, the latest consensus sequences (1996 and 1997) again contained an E at this position. Our subsequent analysis of the amino acid distribution at this position among individual sequences revealed that both early and late sequences had various amino acids at this position [1988, 25 E, two K; 1992, one E, four K, one R (arginine); 1995, seven E, 10 K, three R; 1996, 12 E, five K, one R, one N; 1997, three E, one K, two R].

To test whether the 1984 and 1985 V3 sequences are close to the founder main C virus which was originally introduced into Ethiopia, we reconstructed the most recent common ancestor of the Ethiopian main C virus epidemic as the common node (Fig. 2) of 81 Ethiopian

```
                1        10        20        30        40        50        60        70        80
                |        |         |         |         |         |         |         |         |
CONSENS_84-97   IIIRSENLTNNAKIIIVQLNESVEINCTRPNNNTRESIRIGPGQTFYATGDIIGDIRQAHCNISGEKWNKTLQKVKEKLKEHFPNKTI
1984____84      ...................K.P...............................................T...?.R...E.....R..
1985____85      .......................?.E..................A....K...........T....D......G.?R......-....
C_E2220_86      T...F....................T....S.................................E..............QK.......
CONSENS_84-86   .........................?...............................................?.?...?........
CONSENS_88      ...........................................................................R....?.......
CONSENS_92      .............T.......?.............K............................E...?TV....E?..E........
CONSENS_95      .....................T.............K....................................................
CONSENS_96      .?..............................................................?........?..............
CONSENS_97      ................................................................R........R..............
```

Fig. 1. Predicted amino acid sequences of the *env* gp120 V3 region of the Ethiopian 1984 and 1985 samples. The sequences are shown in comparison with the Ethiopian main C consensus sequences. Dots indicate amino acid identities and dashes indicate deletions. Question marks represent cases when two or more amino acids were the most often ones at that position. The consensus sequences for several periods of the Ethiopian epidemic are also shown.

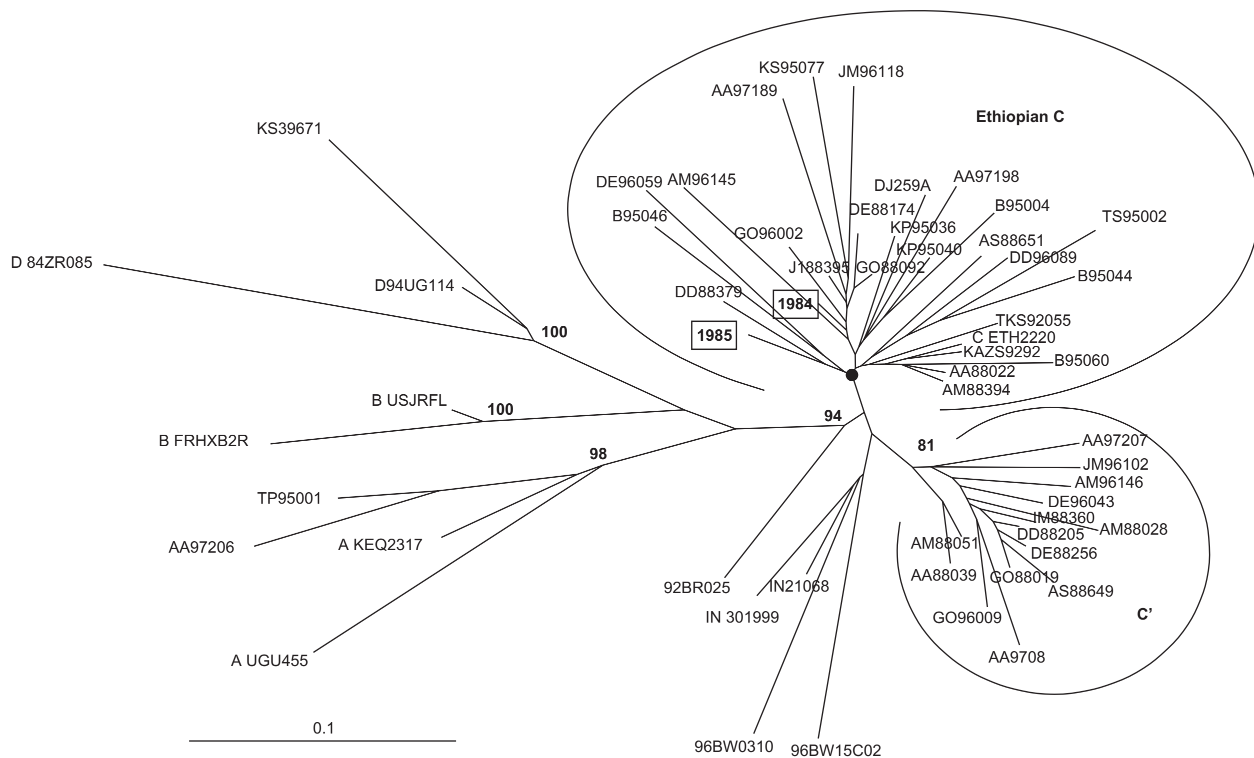

Fig. 2. Neighbor-joining phylogenetic tree of the Ethiopian 1984 and 1985 V3 sequences. Previously obtained Ethiopian main C group and subcluster C′ sequences, as well as sequences of HIV-1 subtypes other than C and subtype C sequences from India and Brazil, are also included. All reference sequences are labelled as in the original studies. The 1984 and 1985 sequences are boxed. The reconstructed most recent common ancestor of the Ethiopian main C virus epidemic is marked by a closed circle. The main C group and subcluster C′ are indicated. Bootstrap values > 80% are shown (percentage of 100 replications). In the *gag* and *pol* regions, the sequences of the 1984 and 1985 samples also belonged to the main C group (not shown).

main C group V3 sequences with known sampling years obtained during 1984–1997. This genomic region was used because intensive longitudinal sequence data for the V3 region are available from Ethiopia. Subsequent analysis of the evolutionary distances of the 1984 and 1985 sequences to the reconstructed common node of the Ethiopian main C virus epidemic revealed that these sequences are closer to the founder virus that are the later sequences. The difference between the distances of the 1984–1985 and later samples to the reconstructed founder virus was especially apparent when only synonymous substitutions were analysed, with both the 1984 and 1985 samples (100%) having a zero D_s to the node. Among later samples (1986–1997), only three out of 79 sequences (4%) had zero D_s to the node. All three later samples with zero D_s to the node were obtained in 1988 (three out of 27 sequences, 11%), whereas none of 51 sequences (0%) from 1992–1997 had a zero D_s to the reconstructed founder virus. The mean D_s (± SE) of later samples to the node were: 1988, 0.040 ± 0.007; 1992, 0.063 ± 0.013; 1995, 0.094 ± 0.012; 1996, 0.094 ± 0.013; 1997, 0.098 ± 0.030. The mean nucleotide distance of the 1984–1985 sequences to the node calculated by using the neighbour-joining method was also lower than that of samples taken during any of later periods: 1984–1985, 0.040 ± 0.024; 1986, 0.049 (one sample, SE cannot be calculated); 1988, 0.063 ± 0.005; 1992, 0.088 ± 0.018; 1995, 0.112 ± 0.009; 1996, 0.093 ± 0.007; 1997, 0.115 ± 0.011 (1986–1997, 0.090 ± 0.043). Similar results were obtained by using the maximum-likelihood method (data not shown).

Regression analysis of evolutionary distances of 81 Ethiopian main C V3 sequences revealed that there is a highly significant ($P < 0.0000001$) positive correlation between sampling years of individual sequences and their D_s to the reconstructed common ancestor of the Ethiopian main C virus epidemic (Fig. 3). The extrapolation of the regression line of D_s back to the date when no synonymous heterogeneity was present in the Ethiopian HIV-1 main C virus population allowed us to estimate 1983 (95% CI, 1980–1984) as the year of HIV-1 main C virus introduction into Ethiopia. Similar results were obtained when regression analysis was based on the mean D_s of all sequences sampled in each year (Fig. 3). This approach resulted in a marked increase of correlation (r, 0.99; r^2, 0.97) between the mean D_s of sequences and their sampling years.

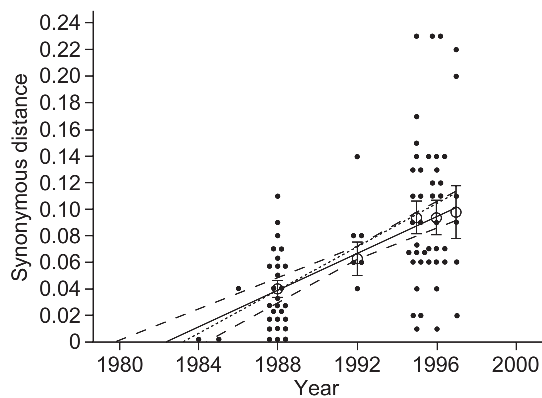

Fig. 3. Synonymous distances of the Ethiopian main C virus V3 sequences to the reconstructed common ancestor of the Ethiopian main C virus epidemic in relation to sequence sampling years. The regression analysis is performed for all 81 individual sequences (closed circles, solid regression line) as well as for the mean D_s per sampling year (open circles, dotted regression line; error bars are shown). Dashed lines indicate 95% CI calculated based on 81 individual sequences. For individual sequences: r, 0.56; r^2, 0.31; slope, 0.008; x-intercept, 1983; $P < 0.0000001$. For the mean D_s: r, 0.99; r^2, 0.97; slope, 0.007; x-intercept, 1983; $P < 0.001$.

Discussion

In the present study, we were able to demonstrate and further confirm a very low prevalence of HIV-1 in Addis Ababa, Ethiopia, in the early 1980s. Among the 474 samples from 1982–1985 tested in our study, only two were found to be positive for anti-HIV-1 antibodies. None of the 98 samples from 1982 and 1983, obtained from HBV infected patients in Addis Ababa, were HIV-1 positive. Only in 1984 was the first HIV-1 positive sample identified, resulting in a low prevalence of 0.5% (1/193 samples). This prevalence remained unchanged in 1985, when one of 183 samples (0.5%) was found to be HIV-1 positive. Our data are in agreement with previously published data on the prevalence of HIV-1 in Ethiopia in the early 1980s [1,2]. When taken together with HIV-1 prevalence of up to 38% among CSW in 23 cities in Ethiopia in 1988 [5], the above data indicate an extremely rapid spread of HIV-1 infection through the country between 1984 and 1988.

From both the 1984 and 1985 HIV-1 positive samples, we were able to obtain genetic information for the *env* gp120 V3 and *pol* or *gag* regions of the HIV-1 genome. To the best of our knowledge, these two samples represent the earliest sequences available for the Ethiopian epidemic. Phylogenetic analysis revealed that both samples belonged to the Ethiopian main C group and not to subcluster C′.

To study virus diversification within the Ethiopian C epidemic, we analysed a set of 81 V3 sequences obtained during 1984–1997 for this and earlier studies. We were able to demonstrate that the amino acid consensus sequence of the Ethiopian HIV-1 main C virus population has remained stationary in the course of the epidemic. Although amino acid variations between consensus sequences of different periods of the epidemic were observed, none of these differences was consistent (Fig. 1). In other words, none of the amino acids that were absent or in minority among early Ethiopian viruses has subsequently been selected for during the course of the epidemic. This observation is similar to earlier findings for the subtype B epidemic [29,30]. Together with increasing genetic heterogeneity among individual sequences over time, which we showed in this study, this observation demonstrates that evolution of HIV-1 main C virus population in Ethiopia is characterized by an unbiased expansion around a stationary consensus sequence.

To test whether the 1984 and 1985 V3 sequences are close to the virus which had originally been introduced into Ethiopia, we used several phylogenetic methods to reconstruct the most recent common ancestor of the Ethiopian main C virus epidemic. Our comparison of evolutionary distances of the 1984 and 1985 sequences with those of sequences obtained later in the epidemic revealed that the 1984–1985 sequences are closer to the ancestor. While being evident for nucleotide distances, this difference was especially pronounced for D_s, with none of the 1984 and 1985 sequences, compared with 89% of the 1988 and 100% of the 1992–1997 sequences, having synonymous substitutions related to the ancestor. This observation agrees with previous data for other epidemics which indicate that (synonymous) genetic heterogeneity of virus populations is increasing in the course of the epidemic [29–33].

By analysing the D_s of the Ethiopian sequences to the ancestor during the course of the epidemic (1984–1997), as was carried out for the US and Dutch HIV-1 epidemics [34], we were able to demonstrate a highly significant correlation with sequence sampling years (Fig. 3). Extrapolation of the regression line back to the date when no synonymous heterogeneity was present in the Ethiopian HIV-1 main C virus population allowed us to estimate 1983 (95% CI, 1980–1984) as the year of HIV-1 C introduction into Ethiopia (Fig. 3), in agreement with seroepidemiological data (Table 1).

Acknowledgements

This study is part of the Ethiopian–Netherlands AIDS Research Project, a collaborative effort of the Ethiopian

Health and Nutrition Research Institute, the Amsterdam Municipal Health Service, the Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, and the Academic Medical Center of the University of Amsterdam. The Ethiopian–Netherlands AIDS Research Project ENARP is supported financially by the Netherlands Ministry of Foreign Affairs and the Ethiopian Ministry of Health.

References

1. Zewdie D. **AIDS Control Programme and Research in Ethiopia.** *24th Ethiopian Annual Medical Conference.* Addis Ababa, May 1988 [abstract 75].
2. Tsega E, Mengesha B, Nordenfelt E, Hansson BG, Lindberg J. **Serological survey of human immunodeficiency virus infection in Ethiopia.** *Ethiop Med J* 1988, **26**:179–184.
3. Lester FT, Ayehunie S, Zewdie D. **Acquired immunodeficiency syndrome: seven cases in an Addis Ababa hospital.** *Ethiop Med J* 1988, **26**:139–145.
4. Mehret M, Khodakevich L, Zewdie D, *et al.* **HIV-1 infection among employees of the Ethiopian Freight Transport Corporation.** *Ethiop J Health Dev* 1990, **4**:177–182.
5. Mehret M, Khodakevich L, Zewdie D, *et al.* **HIV infection and related risk factors among female sex workers in the urban areas of Ethiopia.** *Ethiop J Health Dev* 1990, **4**:163–170.
6. National AIDS Program. *Review Report.* Addis Ababa, Ethiopia: Ministry of Health; 1994.
7. Fontanet AL, Messele T, Dejene A, *et al.* **Age- and sex-specific HIV-1 prevalence in the urban community setting of Addis Ababa, Ethiopia.** *AIDS* 1998, **12**:315–322.
8. Aklilu M, Messele T, Tsegaye A, *et al.* **Factors associated with HIV-1 infection among sex workers of Addis Ababa, Ethiopia.** *AIDS* 2001, **15**:87–96.
9. Abebe A., Kuiken CL, Goudsmit J, *et al.* **HIV type 1 subtype C in Addis Ababa, Ethiopia.** *AIDS Res Hum Retroviruses* 1997, **13**:1071–1075.
10. Ayehunie S, Johansson B, Salminen M, *et al.* **HIV-1 in Ethiopia: phylogenetic divergence from other HIV-1 strains.** *Virus Gen* 1991, **5**:359–366.
11. Ayehunie S, Johansson B, Sonnerborg A, Zewdie D, Britton S, Strannegard O. **Sequence analysis of selected regions of the env (V3 loop and gp41) and gag (p7) reading frames of Ethiopian human immunodeficiency virus type 1 strains.** *Arch Virol* 1993, **128**:229–239.
12. Sherefa K, Sonnerborg A, Steinbergs J, Sallberg M. **Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping and its relation to sequence analysis.** *Biochem Biophys Res Commun* 1994, **205**:1658–1664.
13. Sherefa K, Sallberg M, Sonnerborg A. **Evidence of no change in V3 loop antibody recognition pattern in HIV type 1 infected Ethiopians between 1988 and 1993.** *AIDS Res Hum Retroviruses* 1994, **10**:1551–1556.
14. Sherefa K, Sallberg M, Johansson B, Salminen M, Sonnerborg A. **Subtyping of human immunodeficiency virus type 1 strains by using antibodies specific for the third variable domain (V3) of gp120: results may be affected by divergent V3 sequences.** *J Clin Microbiol* 1997, **35**:2419–2421.
15. Salminen MO, Johansson B, Sonnerborg A, *et al.* **Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C.** *AIDS Res Hum Retroviruses* 1996, **12**:1329–1339.
16. Abebe A, Demissie D, Goudsmit J, *et al.* **HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS.** *AIDS* 1999, **13**:1305–1311.
17. Hussein M, Abebe A, Pollakis G, *et al.* **HIV-1 subtype C in commercial sex workers in Addis Ababa, Ethiopia.** *J Acquir Immune Defic Syndr* 2000, **23**:120–127.
18. Sonnerborg A, Sherefa K. **HIV-1 subtype C epidemic in Ethiopia [letter].** *AIDS* 1997, **11**:1897–1897.
19. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM. **Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients.** *AIDS Res Hum Retroviruses* 1999, **15**:647–653.
20. Abebe A, Pollakis G, Fontanet AL, *et al.* **Identification of a genetic subcluster of HIV type 1 subtype C (C′) widespread in Ethiopia.** *AIDS Res Hum Retroviruses* 2000, **16**:1909–1914.
21. Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, van der Noordaa J. **Rapid and simple method for purification of nucleic acids.** *J Clin Microbiol* 1990, **28**: 495–503.
22. de Wolf F, Hogervorst E, Goudsmit J, *et al.* **Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics.** *AIDS Res Hum Retroviruses* 1994, **10**:1387–1400.
23. Lukashov VV, Cornelissen MJ, Goudsmit J, *et al.* **Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania.** *AIDS* 1995, **9**:435–439.
24. Lukashov VV, Kuiken CL, Goudsmit J. **Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period.** *J Virol* 1995, **69**:6911–6916.
25. Cornelissen M, Kampinga G, Zorgdrager F, Goudsmit J, and the UNAIDS network for HIV isolation and characterization. **Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes**. *J Virol* 1996, **70**:8209–8212.
26. de Baar MP, van der Schoot AM, Goudsmit J, *et al.* **Design and evaluation of a human immunodeficiency virus type 1 RNA assay using nucleic acid sequence-based amplification technology able to quantify both group M and O viruses by using the long terminal repeat as target.** *J Clin Microbiol* 1999, **37**: 1813–1818.
27. Cornelissen M, van den Burg R, Zorgdrager F, Lukashov VV, Goudsmit J. ***Pol* gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D.** *J Virol* 1997, **71**:6348–6358.
28. Kumar S, Tamura K., Nei M. *Molecular Evolutionary Genetics Analysis (MEGA). Version 1.01.* Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania, USA. 1993.
29. Kuiken CL, Lukashov VV, Baan E, Dekker J, Leunissen JAM, Goudsmit J. **Evidence for limited within-person evolution of the V3 domain of the HIV-1 envelope in the Amsterdam population.** *AIDS* 1996, **10**:31–37.
30. Lukashov VV, Goudsmit J. **Evolution of the human immunodeficiency virus type 1 subtype-specific V3 domain is confined to a sequence space with a fixed distance to the subtype consensus.** *J Virol* 1997, **71**:6332–6338.
31. Kuiken CL, Zwart G, Baan E, Coutinho RA, van den Hoek JAR, Goudsmit J. **Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic.** *Proc Natl Acad Sci USA* 1993, **90**:9061–9065.
32. Korber B, Theiler J, Wolinsky S. **Limitations of a molecular clock applied to considerations of the origin of HIV-1.** *Science* 1998, **280**:1868–1871.
33. Goudsmit J, Lukashov VV. **Dating the origin of HIV-1 subtypes.** *Nature* 1999, **400**:325–326.
34. Lukashov VV, Goudsmit J. **HIV-1 intrasubtype diversification originates from the moment of introduction of virulent HIV-1 strains into current AIDS risk groups in the 1960s and 1970s.** *XII World AIDS Conference.* Geneva June–July 1998 [abstract 11197].