T. de O., O. G. P. and A. R. contributed equally to this analysis.

Benghazi Study Group: Carla Montesano (Department of Biology, University TorVergata, Rome, Italy), Francesca Ceccherini-Silberstein (Department of Biochemical Sciences, University Tor Vergata, Rome, Italy), Stefania Montieri (Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy), Danail Beskov (Institute of Infectious and Parasitic Diseases, Sofia, Bulgaria), Sabrine Yerly (University Hospital of Geneva, Switzerland), David Pauza (Institute of Human Virology, University of Maryland, US), Luc Montagnier (World Foundation for AIDS Research and Prevention, Paris, France).

SUPPLEMENTARY INFORMATION

1. Supplementary Notes:

Background of HIV-1 Outbreak at Al-Fateh Hospital:

In May 1998, the Al-Fateh Children's Hospital (AFH) in Benghazi, Libya[1] noted their first case of HIV-1 infection. In September 1998, another 111 children who had been admitted to the hospital were found to be HIV-1 positive[1]. The outbreak was reported by local hospital authorities and representatives from the World Health Organization (WHO) were sent to AFH in December 1998 to examine the cause of the infections[2]. The resultant WHO report suggests that there were multiple nosocomial HIV-1 infections at AFH. The report also notes the lack of required medical equipment in the hospital[2].

Additional Information on the HIV-1 and HCV infected Children:

In total 418 children were infected with HIV-1 in the AFH outbreak. Following the WHO report, 248 (59.3%) of these children were sent to hospitals in Geneva, Rome and Milan for care, treatment and virological assessment[1]. Epidemiological data, available for 37 of the children[1], indicated that all of these children had undergone invasive procedures while in the hospital, or as outpatients. The mean number of hospital visits between January 1998 to April 1999, was 2.14 (range. 1-4) for outpatients; 1.56 (range, 1-6) for hospitalized patients[1]. The median age among the entire group of 248 children, at the time of first diagnosis, was 3.9 ± 4.1 years. The median $\log_{10}$ viral load was 4.61 (range, 1.4-7.8) RNA copies/mL plasma, and median CD4+ percentage was 28% (range, 7-54%). At the time of first observation, 216

(87.1%) of the children were asymptomatic, or mildly symptomatic; 29 (11.7%) had moderately severe symptoms, and 3 (1.2%) had severe symptoms according the CDC classification system. Serological testing of a subset of plasma specimens indicated that 75 (43.1%) of 174 children were co-infected with HCV. Supplementary Table 1 gives more information on clinical aspects of infected children.

Accusation of Foreign Medics:

In March 1998 six foreign medics (five Bulgarian nurses and a doctor from Palestine) joined the medical staff at AFH. One year later, these individuals were accused of purposefully infecting more than 400 children with HIV-1[3]. They have been detained in prison ever since. In April 2003, at the court's request, two international HIV/AIDS scientists, Luc Montagnier and Vittorio Colizzi conducted a scientific inquiry into the Benghazi outbreak[2]. In their report, they conclude that given the high rate of Hepatitis B and C infection amongst the children, the contamination was more likely to be caused by pre-existing poor hygiene practices rather than a single introduction[2]. Furthermore, their report suggests that the HIV-1 and HCV epidemics were present in the hospital prior to the arrival of the Bulgarian medical staff[2]. However, the Libyan court found this report to be imprecise and lacking in evidence and therefore decided not to consider its findings in the trial[4]. In December 2003, a second scientific report produced by Libyan researchers was written for the court[5]. This document has been central to the prosecutor's case against the six medics

In May 2004, the foreign medical staff were condemned to death. However, in response to international appeal, the Libyan Supreme Court ordered a retrial on the 25th of December 2005. The new trial began in Tripoli on the 11th of May 2006, and on the 29th of August, the prosecution again called for the medics to be sentenced to death[4]. The last session of the trial began on the 4th of November 2006 and the final verdict is due to be given on the 19th of December. Attorneys from *Lawyers without Borders,* who are representing the defendants, have appealed to international AIDS experts to conduct an independent scientific inquiry into the history of the Benghazi HIV-1 outbreak[4]. This paper is a response to their appeal.

2. Supplementary Methods:

Study Populations:

The demographic and clinical data for the subset of 44 children available for HIV-1 sequence analysis was comparable to that of the entire European cohort. The median age, viral load and % CD4+ T-cell counts for this subset were 5.9 years (range, 2.2-17.9), 4.49 $\log_{10}$ RNA copies/mL and 27.9% (range, 8-45%), respectively. Thirty-three (75.0%) of these children were asymptomatic, or mildly asymptomatic (N/A), 11 (25%) had moderately severe symptoms, and 1 (2.3%) had severe symptoms. Twenty one (47.7%) had serological evidence of HCV co-infection. Plasma specimens for HIV-1 analyses were collected from 44 children visiting the Bambino Gesù Children Hospital, Rome, Italy between 2000 and 2003. The median age of children sampled in July-August of 2000 was 6.3 years, compared with a median age of 6.0 for all children sampled over the entire three-year period. Supplementary Table 2 provides information on the sampling dates. A total of 66 plasma, collected from children visiting University Hospital in Geneva between 1998 and 2001, were available for HCV sequence analysis as previously described[1].

RNA Extraction, PCR Amplification and Sequencing

For HIV-1 sequence analysis, RNA was extracted from plasma using the QIAamp Viral RNA kit (Qiagen, Heiden, Germany), reverse transcribed and PCR amplified with SuperScript One-Step RT-PCR for Long Templates (Invitrogen) in a 50 µl reaction containing 25 µl of reaction mix, 8 µl of Mg2SO4, 1.5 µl of RT-TAQ, 0.15 µM of sense and antisense primers, and 40 U of RNAse out (Invitrogen). Conditions for the reaction were: one cycle of RT to 50° C for 30 min, one cycle at 94°C for 2 min., 40 cycles (95°C 30 sec., 53°C 30 sec., 72 °C 2 min) and a final cycle at 72°C 10 min. PCR primers used in the reaction were: sense (543, *gag*) [5'gcc tca ata aag ctt gcc tt 3']; antisense (2400, *pol*) [5'cca att ccc cct atc att ttt 3']. Sequencing of the gag region was performed using a BigDye terminator v. 3.1 cycle sequencing kit (Applied Biosystems) and primer 1 [5' gac tag cgg agg cta gaa 3'], primer 2 [5'ggg gtg gct ccc tct gat aa 3'], primer 3 [5' cca gaa gta ata ccc atg tt 3'], and primer 4 [5' cct gac atg ctg tca tca t 3'].

HCV RNA in plasma was quantified with the Cobas Amplicor Monitor Assay (Roche). For sequence analysis, total RNA was extracted and the hypervariable region 1 of the envelop glycoprotein E2 (nt 952-1365) was amplified using primers specific for genotypes 1 and 4. These primer sets were: sense (952, E2) [5□-atg gcn tgg gay atg atg atg aay tgg-3], antisense (1365, E2) [5□-ccg ytc ggr rca scc tga rgt tra ayt tgt-3□]; and sense (1048, E2) [5□-ggn ngg bca ctg ggg hrt yct-3□], antisense (1265, E2) [5□-atr tgc car cts ccr ttg stg ttga-3□]. Reaction conditions consisted of a 5-min hot start step at 95°C, followed by 35 cycles with a denaturation step at 94°C for 30 s, an annealing step at 55°C for 45 s, an elongation step at 72°C for 45 s, and a final extension step at 72°C for 7 min. Direct sequencing was performed using: primer 1 [5-ggn ngg bca ctg ggg hrt yct-3] and primer 2 [5-atr tgc car cts ccr ttg stg ttga-3].

Reference Sequences and Alignment

A reference dataset for HIV-1 CRF02_AG was obtained from the HIV Sequence Database at Los Alamos[6] and from GenBank. In total, 56 reference strains with the highest BLAST search similarity scores to the AFH strains were chosen (see Supplementary Table 3). These include strains from Cameroon (24), Ivory Coast (4), Ghana (7), Nigeria (3), Niger (2), Senegal (2), Gambia (1), Djibouti (1) and Democratic Republic of Congo (1), USA (4), Sweden (1), France (2), Italy (1) and Uzbekistan (3). All of the reference sequences and all of the AFH HIV-1 sequences were subtyped using the REGA HIV-1 Subtyping Tool[7], version 2.0. A reference dataset for HCV was obtained from the HCV sequence database[8]. Preliminary phylogenetic analysis was used to reduce an initial 12,000 E1/E2 HCV sequences to 210 closely related reference strains (113 genotype 4 sequences and 97 genotype 1 sequences; see Supplementary Table 4). The AFH HIV and HCV sequences were then aligned with their respective reference datasets using ClustalW[9] and subsequently edited by hand using Se-Al 2.0[10].

Phylogenetic Analysis

The best fitting nucleotide substitution model was evaluated using hierarchical likelihood ratio tests – the methodology implemented in ModelTest[11]. The model chosen for the HIV-1 alignment was GTR+G (general time reversible model with

gamma distributed among-site rate heterogeneity). The best fitting model for the HCV alignment was HKY+G+I (Hasegawa Kino Yano model with gamma rate heterogeneity, plus invariant sites). Maximum likelihood (ML) phylogenies were estimated for each dataset under the abovementioned models, starting with a NJ starting tree and using the TBR (tree bisection and reconnection) and SPR (subtree pruning and regrafting) heuristic search algorithms. The final ML phylogenies are shown in Supplementary Figures 1, 2, and 3. Calculations were performed with PAUP* 4.0b10[12]. Statistical support for ML phylogeny structures was evaluated by bootstrapping analysis of the original sequence alignments (1000 NJ replicates). Phylogenetic branches were also investigated using the ML-based zero branch length test implemented in PAUP* 4.0b10[12]. Trees were rooted using mid-point rooting and presented using the program FigTree[13]. Bayesian estimates of phylogeny were obtained using MrBayes[14], under the GTR+G model for the HIV alignment and the HKY+G+I model for the HCV alignment. For each dataset, two Markov Chain Monte Carlo (MCMC) runs were calculated independently. Each MCMC run was 10,000,000 steps long and was sampled every 1000 steps, with a temperature parameter of 0.2. The average standard deviation of split frequencies was calculated for the two chains to check for convergence. Subsequently, the Effective Sampling Size (ESS) was calculated by combining the output of the two runs using Tracer, excluding an initial burn-in of 10% for each chain[15]. The ESS values were >1000, indicating a sufficient level of sampling. A final Bayesian majority-rule consensus tree was obtained for the HIV (see Supplementary Figure 4).

Genome-region specific independent estimates of evolutionary rate

Evolutionary analyses were conducted in order to estimate a real timescale for the AFH infection clusters. The first step in this analysis was to obtain estimates for the evolutionary rate of the sequenced virus genome regions from appropriate datasets; this information was then encoded as a prior probability distribution and combined with the AFH sequence data using an established Bayesian MCMC approach[16] that has been tested in cases where epidemiological history is well known[17]. This is an important step as evolutionary rates vary widely among different HIV and HCV genomic regions[18,19].

The HIV evolutionary rate for the sequenced region was estimated using a set of reference strains obtained from the HIV Sequence Database[6]. To be included in this set, reference strains had to span the entire region of interest (positions 790-1722), belong to the A1 sub-subtype of HIV-1, and have a recorded date of sampling. Based on these criteria, a total of 48 reference sequences, 921 bp in length, with sampling dates ranging from 1985 to 2004, were identified and used in the alignment (see Supplementary Table 4). Evolutionary rates were estimated from this alignment using both a strict and a relaxed molecular clock (uncorrelated lognormal model), as implemented in Beast 1.4[16]. A nucleotide substitution model that takes account of site rate variation due to codon structure[20] was used in both cases. Under the strict clock model, the estimated HIV evolutionary rate was 1.47×10^{-3} substitutions per site per year (95% HPD confidence limits: 1.02×10^{-3} to 1.96×10^{-3}). The relaxed clock model[21] evolutionary rates were similar to the strict clock results; the estimated average HIV evolutionary rate was 1.59×10^{-3} substitutions per site per year (0.9×10^{-3} to 2.3×10^{-3}). The coefficient of variation was estimated at 0.28 (0.18 to 0.38), indicating relatively little variation in evolutionary rate among branches.

To obtain a rate of HCV sequence evolution we utilized sequence data from a previous case of iatrogenic HCV transmission, the Irish anti-D cohort. We used HCV complete genome sequences obtained by direct PCR from 15 patients, 17 to 21 years after the patients were infected with contaminated anti-D immunoglobulin[22] (see Supplementary Table 6). The relevant sub-genomic regions (positions 1408-1580) were extracted from the anti-D cohort HCV complete genomes, and compared with the HCV sequence obtained from the anti-D outbreak source, which contained very little HCV genetic diversity[23]. We followed a similar approach to that used by Pybus and colleagues[24]. The anti-D cohort infection history was represented as a star phylogeny and the HCV evolutionary rate was then estimated using a Bayesian MCMC framework[16] under a strict molecular clock (sequence length was insufficient to support a reliable relaxed clock analysis of this data). The HCV evolutionary rate was estimated at 6.0×10^{-3} substitutions per site per year (95% HPD confidence limits: 5.3×10^{-3} to 6.8×10^{-3}) using the HKY evolutionary model and estimated at 16.7×10^{-3} substitutions per site per year (11.9×10^{-3} to 22.0×10^{-3}) using the HKY+G evolutionary model.

Evolutionary Analysis of the AFH Clusters.

To estimate the age of the individual HIV and HCV AFH clusters, we employed both a constant rate molecular clock and a relaxed clock model[21] under a range of plausible models of epidemiological change and nucleotide substitution. This approach accommodates uncertainty in the parameters of the model including phylogenetic structure.

The estimated marginal posterior probability distributions for the rate from the HIV-1 and HCV calibration datasets were used as a calibration prior to calculate the age of their respective AFH clusters[21]. The estimated values for the AFH HIV-1 sequences varied from 1.5×10^{-3} to 2.0×10^{-3} depending on the combination of tree and molecular clock model applied (Supplementary Tables 7 and 8). The estimated evolution rate of the HIV-1 AFH sequences was similar to rates previously described for this genomic region[19]. The estimated values for the HCV clusters varied from 6.0×10^{-3} to 18.0×10^{-3} depending on which model was used. This is consistent with the estimated rate from the anti-D cohort. The HCV rates were only estimated using strict molecular clock models, since the short size of the clusters and the sequence length do not allow one to use more parameter rich models, such as relax molecular clock model[17]. The marginal posterior estimated rates under the complete range of models applied in this study is included in Supplementary Tables 7 and 8.

We summarized the results by looking at the distribution of ages of the most recent common ancestor of each cluster and determining the posterior probability that this included March 1998. We also calculated the proportion of lineages that existed prior to this date averaged over the sampled phylogenies. Three population genetic models were applied to analyze the data. The first model assesses that the population size is constant over time[25,26]. The second assumes that the population is growing exponentially over time[25,26]. The third, the Bayesian Skyline plot[27] (BSP), determined the population growth model using the supplied data. No matter which model was used, the estimated date of the MRCA of each HIV-1 and HCV cluster predated March 1998 (Supplementary Tables 7 and 8). In most analyses the probability that the AFH clusters originated after March 1998 was practically zero. However, combining both relaxed clock and the Bayesian skyline models, we found a probability of 0.10. But one must take in consideration that this was the most complex model combination, hence the increased confidence intervals obtained probably reflect model over-parameterization (see Supplementary Tables 8) and thus loss of

power. The analyses were performed using 100,000,000 chains with sampling each 10,000 chain using Beast v1.4[16]. For each model, 4 independent chains were run. The results were visualized in Tracer[11] and the convergence and mixing quality of the runs determined. The ESS values were >1000, indicating a sufficient level of sampling. The results reported are the combined estimates of the 4 independent runs.

Eric Delwart kindly provided previously unpublished HCV sequence data from Egypt (Supplementary Table 9), Philippe Lemey and Peter Markov assisted with analysis, and Paul Harvey commented on the manuscript.

3. Supplementary References:

1. Yerly, S., et al. Nosocomial outbreak of multiple blood borne viral infection. J Infect. Dis. **184**, 369-372 (2001).
2. Montagnier L & Colizzi V. "Report on the Benghazi outbreak" Available from http://www.nature.com/nature/journal/v443/n7114/extref/montagnier.pdf (2006).
3. Nature Editorial. A true test of leadership. Nature 430, 277(2004).
4. Butler D. Lawyers call for science to clear AIDS nurses in Libya. Nature 443, 254 (2006).
5. Final Report by The National Experts Committee Regarding The Scientific Expert Opinion (http://www.nature.com/nature/journal/v443/n7114/extref/national_experts.pdf).
6. Los Alamos HIV Sequence Database. http://www.hiv.lanl.gov (2006).
7. De Oliveira, T., et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics **21**, 3797-3800 (2005).
8. Kuiken, C., Yusim, K., Boykin, L., Richardson, R. The Los Alamos HCV Sequence Database. Bioinformatics **21**, 379-384 (2005).
9. Thompson, J.D., Higgins, D.G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680 (1994).
10. Rambaut, A. 1996. Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/. (1996)
11. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. **14**, 817-818 (1998).
12. Swofford, D.L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts (2003).
13. Rambaut, A. FigTree. Available from http://evolve.zoo.ox.ac.uk/software.html (2006).
14. Ronquist, F., Huelsenbeck, J.P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**, 1572-1574 (2003)
15. Rambaut, A., Drummond, A.J. Tracer. Available from http://evolve.zoo.ox.ac.uk/software.html?id=tracer. (2006)

16. Drummond, A.J., Rambaut, A. BEAST v1.3, Available from http://evolve.zoo.ox.ac.uk/beast/ (2006).
17. Pybus, O.G., Drummond, A.J., Nakano, T., Robertson, B.H., Rambaut, A.The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol. **20**, 381-387 (2003).
18. Salemi, M., Vandamme, A.M. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. J Mol Evol. **54**, 62-70 (2002).
19. Lemey. P., Pybus, O.G., Rambaut, A., Drummond, A.J., Robertson, D.L., Roques, P., Worobey, M., Vandamme, A.M.The molecular population genetics of HIV-1 group O.Genetics **167**, 1059-68 (2004).
20. Shapiro, B., Rambaut, A., Drummond, A.J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol **23**, 7-9 (2003).
21. Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A. Relaxed phylogenetics and dating with confidence.. PLoS Biology 4, e88 (2006).
22. Itakura, J., Nagayama, K., Enomoto, N., Hamano, K., Sakamoto, N., Fanning, L.J., Kenny-Walsh, E., Shanahan, F., Watanabe, M.J. Viral load change and sequential evolution of entire hepatitis C virus genome in Irish recipients of single source-contaminated anti-D immunoglobulin. Viral Hepat. **12**, 594-603 (2005).
23. McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P.L., Simmonds, P., Smith, D.B. Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort. J Virol. **72**, 4893-905 (1998).
24. Pybus, O.G., Charleston, M.A., Gupta, S., Rambaut, A., Holmes, E.C., Harvey, P.H. The epidemic behavior of the hepatitis C virus. Science **292**, 2323-2325 (2001).
25. Pybus, O.G., Rambaut, A., Harvey, P.H. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics **155**, 1429-1437 (2000).
26. Pybus, O.G., Rambaut, A. GENIE: estimating demographic history from molecular phylogenies. Bioinformatics **18**, 1404-1405 (2002).
27. Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G. Mol Biol Evol 22, 1185-1192 (2005).

Supplementary Table 1: Demographic and clinical characteristics of the 248 HIV infected children as determined at the time of their first visit to a European clinic

Patient characteristics at first observation			
	Male (%)	**Female** (%)	**Total** (%)
Number of Patients	139 (56)	109 (44)	248
Median age at diagnosis (y)	3.47 $\pm$3.72	4.53 $\pm$4.55	3.9 $\pm$4.1
CDC class: N or A	122 (87.8)	94 (86.2)	216 (87.1)
CDC class: B	16 (11.5)	13 (11.9)	29 (11.7)
CDC class: C	1 (0.7)	2 (1.8)	3 (1.2)
CD4+ $\geq$25% (CDC-1)	90 (64.7)	78 (71.5)	168 (67.7)
CD4+ 15-24% (CDC-2)	33 (23.7)	24 (22.0)	57 (22.9)
CD4+ <15% (CDC-3)	16 (11.5)	7 (6.4)	23 (9.3)
Median CD4% (range)	27 (7-50)	29 (7-54)	28 (7-54)
VL: < 400 copies/ml	15 (11.9)	12 (11.8)	27 (11.8)
VL: 400-9999 copies/ml	21 (16.7)	28 (27.5)	49 (21.5)
VL: 10000-100000 copies/ml	39 (31.0)	31 (30.4)	70 (30.7)
VL: >100000 copies/ml	51 (40.5)	31 (30.4)	82 (36.0)
median VL $\log_{10}$ (range)	4.73 (1.4-7.0)	4.49 (1.4-7.8)	4.61 (1.4-7.8)

CDC class = clinical and immunological status as defined by the Center for Disease Control and Prevention (CDC) paediatric classification (CDC 1994, MMWR 1994;42 RR:12:1-10). The four categories are: N (asymptomatic), A (mildly symptomatic), B (moderately severe symptoms) and C (severe symptoms)
CD4+ T-cells counts, expressed in percentages
VL = Number of copies of HIV-1 RNA per mL of plasma

Supplementary Table 2: HIV-1 infected sequences generated from patients visiting the Bambino Gesù Children Hospital, Rome, Italy.

Patients	Date of Birth (year.month)	Sample Date	Sample date (years)	Age in years when sampled
LB110	1997.2	27/6/2000	2000.49	3.3
LB113	1997.7	4/7/2000	2000.51	2.8
LB114	1996.8	4/7/2000	2000.51	3.7
LB117	1997.5	4/7/2000	2000.51	3.0
LB123	1987.7	6/7/2000	2000.51	15.8
LB126	1998.0	6/7/2000	2000.51	3.5
LB128	1997.9	6/7/2000	2000.51	2.6
LB129	1990.1	6/7/2000	2000.51	10.4
LB102	1996.8	11/7/2000	2000.52	3.7
LB135	1990.1	17/7/2000	2000.54	10.4
LB210	1996.6	20/7/2000	2000.55	4.0
LB214	1995.1	24/7/2000	2000.56	5.5
LB211	1996.7	25/7/2000	2000.56	3.9
LB212	1992.5	25/7/2000	2000.56	8.1
LB217	1996.5	25/7/2000	2000.56	4.1
LB219	1996.7	25/7/2000	2000.56	3.9
LB139	1989.9	26/7/2000	2000.57	10.7
LB221	1982.7	27/7/2000	2000.57	17.9
LB228	1991.6	27/7/2000	2000.57	9.1
LB233	1997.3	1/8/2000	2000.58	2.7
LB235	1997.7	1/8/2000	2000.58	2.9
LB240	1995.8	1/8/2000	2000.58	4.8
LB310	1997.4	1/8/2000	2000.58	3.2
LB301	1998.0	3/8/2000	2000.59	2.6
LB304	1997.7	3/8/2000	2000.59	2.9
LB306	1992.2	3/8/2000	2000.59	7.7
LB308	1988.3	3/8/2000	2000.59	12.3
LB311	1994.1	3/8/2000	2000.59	6.5
LB313	1994.8	8/8/2000	2000.6	5.8
LB315	1997.6	8/8/2000	2000.6	3.0
LB319	1989.4	8/8/2000	2000.6	11.2
LB321	1997.5	8/8/2000	2000.6	3.1
LB322	1987.8	8/8/2000	2000.6	12.8
LB330	1998.1	9/8/2000	2000.6	2.5
LB112	1998.1	9/1/2001	2001.02	2.9
LB218	1996.9	9/1/2001	2001.02	4.1
LB216	1997.8	18/1/2001	2001.05	3.3
LB229	1997.4	25/1/2001	2001.07	3.7
LB333	1997.2	30/1/2001	2001.08	3.9
LB332	1989.4	6/2/2001	2001.1	11.7
LBPL305	1998.3	8/2/2001	2001.11	2.8
LB214PL	1997.2	3/1/2003	2003.08	5.9
LB225PL	N/A	3/1/2003	2003.08	N/A
LB126PL	N/A	31/1/2003	2003.08	N/A

Supplementary Table 3: HIV-1 reference dataset CRF02_AG used in phylogenetic tree constructions. Sequence naming includes country, year of isolation and the sequence name as stored in the Los Alamos HIV Sequence Database[2].

CM.01.01CM_0002BBY, CM.01.01CM_4410HAL, SN.98.MP1211, CM.01.01CM_0005BBY, CM.02.02CM_1677LE, SN.98.MP1213, CM.01.01CM_0008BBY, CM.97.CM52885, GA.x.LBV2310, CM.01.01CM_0074NY, CM.97.CM53658, DJ.91.DJ258, CM.01.01CM_0131NY, CI.x.CI20, CD.84.30620, CM.01.01CM_0158ND,CI.x.CI51, US.98.98US_MSC5007, CM.01.01CM_0191ND, CI.x.CI59, US.x.00US_MSC3083, CM.01.01CM_0925MO, CI.x.IC144, US.x.98US_MSC404, CM.01.01CM_1237NG, GH.x.GHNJ176, US.x.99US_MSC1134, CM.01.01CM_1475MV, GH.x.GHNJ185, SE.94.SE7812, CM.02.02CM_0013BBY, GH.x.GHNJ188, FR.91.DJ2632, CM.02.02CM_0014BBY, GH.x.GHNJ196, FR.91.DJ264, CM.02.02CM_0015BBY, GH.97.97GHAG1, IT.x.IT067, CM.02.02CM_1669LE, UZ.02.02UZ710, CM.02.02CM_1901LE, UZ.02.02UZ0683, CM.02.02CM_1970LE, NG.01.PL0710, UZ.02.02UZ693, CM.02.02CM_2162SA, NG.01.PL0754, CM.02.02CM_2348SA, NG.x.IBNG, CM.02.02CM_4082STN, NE00_gi37496497, CM.99.pBD6_15, NE.97.gi37496481, GH.94.gi9886936, GH.97.gi9886927

Supplementary Table 4: HIV-1 sequences used in the estimation of the evolutionary rate. Sequence naming includes country, year of isolation and the sequence name as stored in the Los Alamos HIV Sequence Database[2].

GH.97.97GHAG1, FR.91.DJ263, FR.91.DJ264, SE.94.SE7812, SN.98.MP1211, CM.99.pBD6_15, DJ.91.DJ258, CD.85.MAL, GH.97.AG2, MM.99.mCSW105, UG.92.UG029, KE.94.Q23_17, SE.94.SE7535, SE.95.SE8603, SE.95.UGSE8131, BW.98.BW2117, BY.97.97BL006, TZ.X97.97TZ03, UA.00.98UA0116, KE.97.ML170, KE.86.ML170, KE.95.ML170, BE.94.VI1197, CM.97.CM53122, KE.86.ML013_106, KE.97.ML013_2, KE.97.ML605_3, KE.97.ML752, UG.92.UG035, UG.90.UG266, RU.03.03RU20_06, EE.01.EE0369, UG.02.02_4879MH, UG.03.03_8003KN, UG.03.03_9256NI, UG.03.03_9270NM, UG.03.03_9366WG, UG.03.03_9410NS, UG.03.03_9538NG, UG.04.04_0217BM, UG.04.04_0535KH, UG.04.04_0566GT, UG.04.04_0567NR, UG.02.TC024604, UG.02.TC025104, KE.00.CQ891776, UG.85.U455, U00000G.92.92UG037.

Supplementary Table 5: Accession number for the HCV reference dataset used in the phylogenetic tree constructions.

AF488371, AY743031, AY936013, AF488372, AY74303, AY936014, AY742960, AY74303, AY936017, AY742962, AY74303, AY936018, AY742964, AY74303, AY936021, AY742966, AY74303, AY936025, AY742968, AY74304, AY936026, AY742969, AY74304, AY936029, AY742970, AY74304, AY936033, AY742971, AY74304, AY936038, AY742972, AY74304, AY936039, AY742973, AY74304, AY936040, AY742974, AY74304, AY936044, AY742975, AY74304, AY936046, AY742976, AY9359, AY936047, AY742977, AY9360, AY936049, AY742978, AY9360, AY936050, AY742979, AY93601, AY936051, AY742980, AY9360, AY936053, AY742981, AY9360, AY936055, AY742982, AY9360, AY936056, AY742983, AY9360, AY936067, AY742984, AY9360, AY936068, AY742986, AY9360, AY936071, AY742987, AY93604, AY936073, AY742989, AY9360, AY936079, AY742990, AY93606, AY936081, AY742992, AY93606, AY936083, AY742994, AY936082 AY936087, AY742995, AY936095 AY936088, AY742996, AY93610, AY936089, AY742997, AY93611, AY936091, AY742998, AY93611, AY936094, AY742999, AY93611, AY936099, AY743000, AY93611, AY936100, AY743001, AY93611, AY936101, AY743002, AY93612, AY936102, AY743003, AY93612, AY936104, AY743004, AY93612, AY936105, AY743005, AY93613, AY936106, AY743007, AY93613, AY936127, AY743008, D43678, AY236366, AY743009, 43680, AY236389, AY743013, D45193, DQ504442, AY743014, Y11604, AB079081, AY743015, AB079082, AY743016, AF118587, AY743017, AF245810, AY743018, AF268580, AY743019, AF344981, AY743020, AF426595, AY743021, AJ510935, AY743022, AF488358, AJ560356, AY743023, D16189, AJ560418, AY743024, D16191, AY190846, AY743026, AY051292, AY450720, AY743027, X76414, DQ651144, AY743028, AY936002, U14233, AY743029, AY936010, U14230, AY743030, AY936012, U14235, U14238, AB107944, M74808, M74811, AF163248, M74888, AF011753, AF422462, M86769, DQ505105, AJ310591, M62321, AF040777, AJ866094, AY746685, AF054247, DQ508441, AF547456, AF054250, DQ650930, DQ508448, DQ651166, U45476, DQ508457, AY940619, M74804, DQ508458, AB008443, M74805, DQ508460, 4a_Egypt2, 4a_Egypt3, 4a_Egypt4, 4a_Egypt7, 4a_Egypt9, 4a_Egypt10, 4a_Egypt12.

Supplementary Table 6: HCV sequences used in the estimation of the evolutionary rate. The estimated rate was used as a prior to the calculation of the HCV sequences from AFH.

AF313916, AF056733, AF056734, AF056735, AF056736, AF056737, AF056738, AF056739, AF056740, AF056741, AF056742, AF056743, AF056744, AF056745, AF056746, AF056747, AF056748, AF056749, AF056750, AF056751, AF056752, AF056753, AF056754, AF056755, AF056756, AB154177, AB154179, AB154181, AB154183, AB154185, AB154187, AB154189, AB154191, AB154193, AB154195, AB154197, AB154199, AB154201, AB154203, AB154205.

Supplementary Table 7: Evolutionary Sequence Analysis Results

	Tree model	*Sequence evolution model*	*Clock model*	*Rate of evolution**	*Date of MRCA of cluster*	*Percentage of cluster lineages that predate 1/3/98*	*Probability that MRCA of cluster postdates 1/3/98*
HCV cluster 1	Const	HKY+G	strict	0.013 (0.007,0.018)	1992.5 (1988.4,1995.6)	77.1 (68.2,86.4)	<0.001
(n=22)	Expo	HKY+G	strict	0.012 (0.007,0.017)	1994.7 (1992.4,1996.5)	84.0 (77.3,86.4)	<0.001
HCV cluster 2	Const	HKY+G	strict	0.017 (0.011,0.022)	1996.5 (1994.8,1997.7)	62.9 (44.4,77.8)	<0.001
(n=9)	Expo	HKY+G	strict	0.018 (0.012,0.023)	1997.0 (1995.8,1998.0)	63.4 (44.4,77.8)	<0.001
HCV cluster 3	Const	HKY+G	strict	0.011 (0.005,0.017)	1989.2 (1980.2,1995.7)	66.3 (45.8,79.2)	<0.001
(n=24)	Expo	HKY+G	strict	0.012 (0.007,0.017)	1994.3 (1990.7,1997.3)	70.1 (54.2,83.3)	<0.001
HIV cluster	Const	SRD06	strict	0.0018 (0.0012,0.0024)	1994.4 (1991.0,1997.3)	34.7 (15.9,56.8)	<0.001
(n=44)	Expo	SRD06	strict	0.0017 (0.0012,0.0023)	1996.5 (1994.8,1998.1)	42.6 (9.1, 81.8)	<0.005

* Units are nucleotide substitutions per site per year
95% HPD confidence limits are shown in parenthesis
HKY = Hasegawa-Kishino-Yano model (1985); HKY+G = HKY model with gamma distributed among-site rate variation
SRD06 = One HKY+G model for codon positions 1 & 2, another HKY+G model for codon position 3
Const = constant size; Expo = exponential growth; BSP = Bayesian Skyline Plot
Two sequences with no sampling dates could not be included in the HCV cluster 2 analyses

Supplementary Table 8: Extra Models Analysis Results

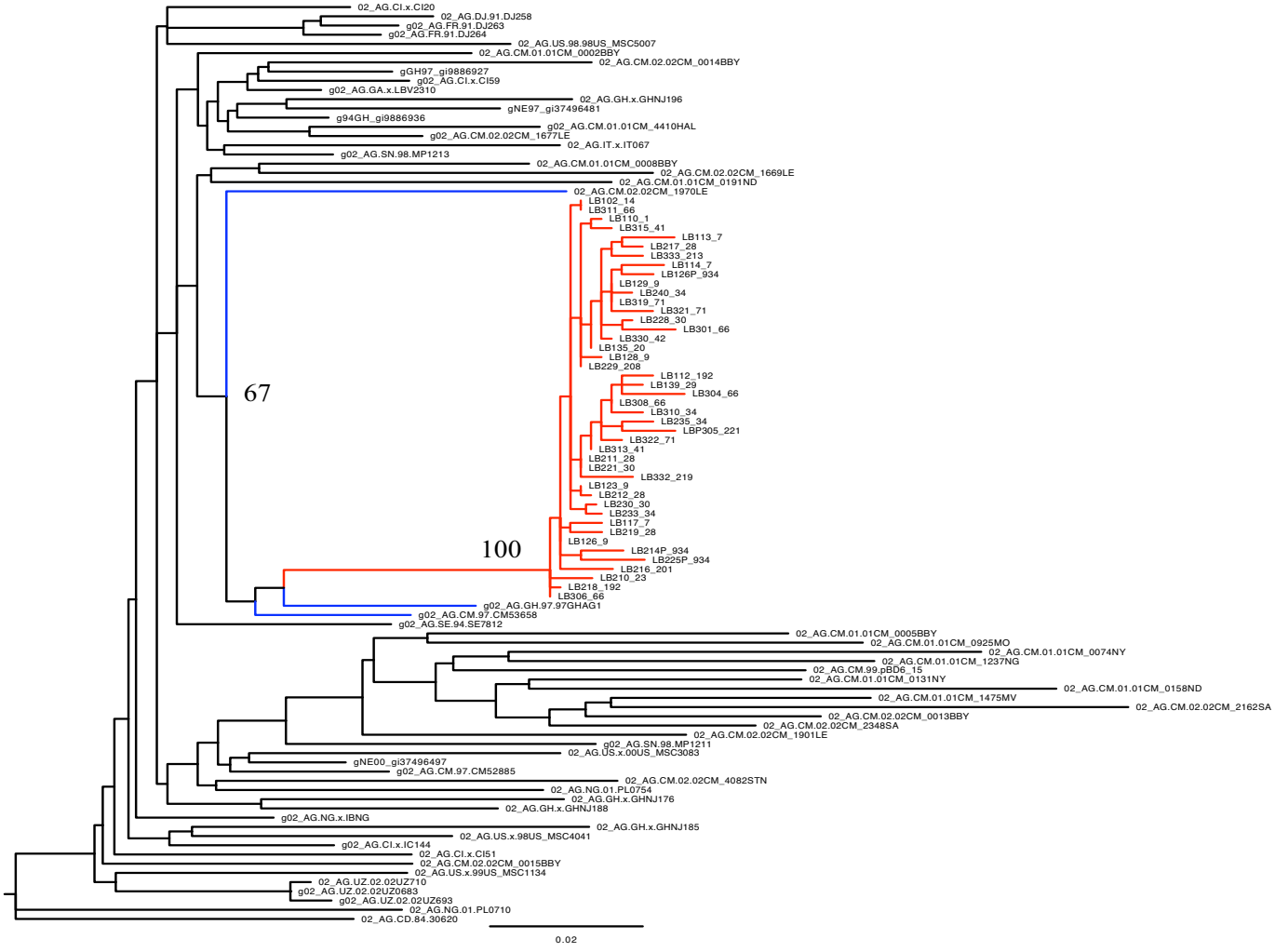
	Tree model	*Sequence evolution model*	*Clock model*	*Rate of evolution**	*Date of MRCA of cluster*	*Percentage of cluster lineages that predate 1/3/98*	*Probability that MRCA of cluster postdates 1/3/98*
HCV cluster 1	const	HKY	strict	0.006 (0.005,0.007)	1989.5 (1986.0,1992.8)	82.9 (77.3,86.4)	<0.001
(n=22)	expo	HKY	strict	0.006 (0.005,0.007)	1992.7 (1990.6,1994.7)	85.6 (81.8,86.4)	<0.001
HCV cluster 2	const	HKY	strict	0.006 (0.005,0.007)	1990.3 (1986.5,1993.8)	83.6 (66.7,88.9)	<0.001
(n=9)	expo	HKY	strict	0.006 (0.005,0.007)	1991.8 (1988.2,1994.7)	85.4 (77.8,88.9)	<0.001
HCV cluster 3	const	HKY	strict	0.006 (0.005,0.007)	1990.3 (1986.7,1993.7)	71.3 (62.5,83.3)	<0.001
(n=24)	expo	HKY	strict	0.006 (0.005,0.007)	1992.9 (1990.1,1995.8)	77.9 (66.7,87.5)	<0.001
HIV cluster	const	SRD06	relaxed	0.0015 (0.0005,0.0025)	1985.5 (1963.2,1997.8)	37.0 (18.2,61.4)	<0.001
(n=44)	expo	SRD06	relaxed	0.0019 (0.0010,0.0027)	1996.6 (1994.4,1998.4)	38.7 (4.5, 88.6)	0.02
	BSP	SRD06	strict	0.0018 (0.0012,0.0024)	1996.7 (1994.7,1998.4)	39.1 (4.5, 84.1)	0.02
	BSP	SRD06	relaxed	0.0020 (0.0011,0.0030)	1996.9 (1993.9,1998.9)	29.3 (2.3, 86.4)	0.10

* Units are nucleotide substitutions per site per year
95% HPD confidence limits are shown in parenthesis
HKY = Hasegawa-Kishino-Yano model (1985); HKY+G = HKY model with gamma distributed among-site rate variation
SRD06 = One HKY+G model for codon positions 1 & 2, another HKY+G model for codon position 3 (ref)
Const = constant size; Expo = exponential growth; BSP = Bayesian Skyline Plot
Two sequences with no sampling dates could not be included in the HCV cluster 2 analys

Supplementary Table 9: Unpublished Egyptian HCV sequences, which were kindly provided by Eric Delwart, spanning a 177 bp fragment of the hypervariable E1/E2 region of the virus.

>4a_Egypt2
CGTGGGAGTAGCTTATTTCAGCATGCAAGCTAATTGGGCCAAAGTCATCT
TAGTCCTATTCCTCTTTGCAGGGGTTGACGCTGAGACTCATGTGTCTGGGG
GTGCGGTTGGCCGAACCGCCCAAGGCCTGACCRGCCTCTTCAGCCCTGGA
GCCCAGCAAACCTTGCAGCTCG
>4a_Egypt3
GGTGGGATTGGCCTACTTCAGCATGCAGGCTAATTGGGCCAAAGTCATCC
TGGTCCTATTCCTCTTTGCAGGGGTGGATGCCGAAACCTATGTGACTGGAG
CGGCAGTTGGTCGCCARGCCGCCAGCTTCACTGGCCTCTTCCAGCATGGGT
CTAGGCAAAACGTGCAGCTCA
>4a_Egypt4
CGTGGGGGTGGCCTACTACTCCATGCAAGCCAATTGGGCCAAAGTCATCC
TAGTCTTATTCCTCTTTGCAGGGGTTGACGCTGAGACTTACACATCTGGGG
GTGCGGCCGCCCAAACTACCCGTGGCTTGGTTAGCCTATTTGGCCCTGGAC
CTCAACAAAAATTGCAGCTCA
>4a_Egypt7
CGTGGGRGTGGCCTATTTCAGCATGCARGCTAATTGGGCAAARGTCATCTT
AGTCCTATTCCTCTTCGCWGGGGTTGACGCTGACACYCRCGTATCYGGGGG
GTKYGGCTGGTYAYAMCCTCMRTGGGKYSRWWRGCMTCTTYTCCCSCGG
AKCYCRGCAAAAWKTGCAGCTCA
>4a_Egypt9
CGTGGGAGTGGCCTATTACAGCATGCAAGCCAATTGGGCCAAAGTCATCT
TAGTCCTGTTCCTTTTTGCAGGGGTTGATGCCAGCACCTACACGACCGGGG
GGGTGGCTGGCAGAGGCGCCAGCCAACTCACTAGTCTCTTCACCGCTGGA
TCTGCGCAGAACTTGCAGCTCA
>4a_Egypt10
CGTRGGATTGGCCTACTTCATCATGCAAGCCAATTGGGCCAAAGYCATCT
TAGTCCTATTCCTCTTYGCAGGRGTTGACGCTGAGACTCACRTATCTGGGG
GTACGTATGGCCGRAACATCCAYGGCCTCACCAGCCTTTTCAGCCCTGGA
TCTCAGCAAAGGTTGCAGCTCR
>4a_Egypt12
CGTAGGAGTGGCCTACTTCTGCATGCAAGCTAATTGGGCCAAAGTCATCT
TAGTCCTATTCCTCTTTGCAGGGGTCGACGCTAACACCTATACGWCTGGG
GGTGCGGCTGGCAGAACCACCCAGGGCCTGACYAGCCTCTTYGCCCCCGG
ATCCCAGCAAAACGTGCAGCTCR

Supplementary Figure 1

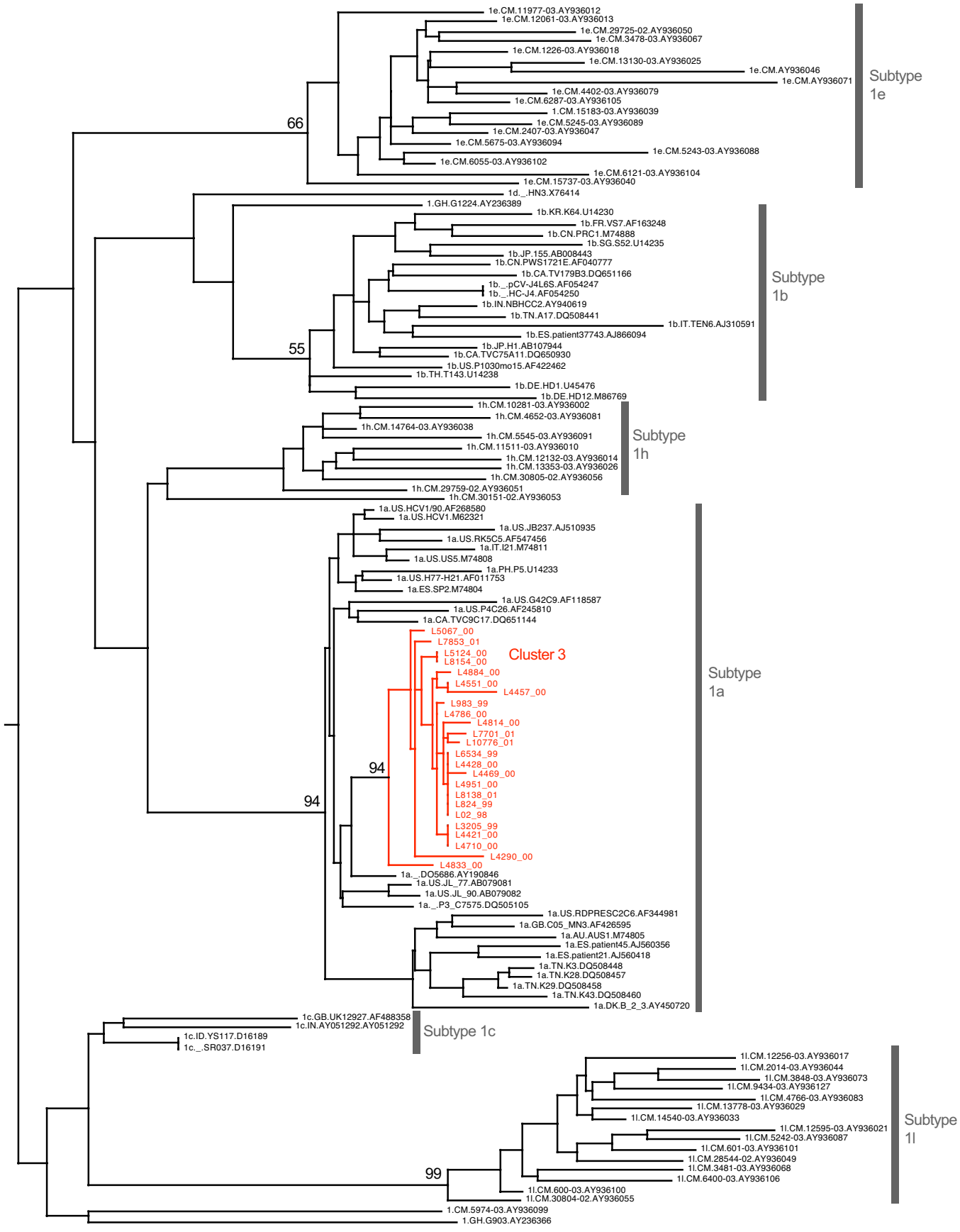


Maximum likelihood phylogenetic tree showing the relationship among HIV-1 CRF02_AG sequences from Libyan (n=44), African (n=46), European (n=4), American (n=4) and Uzbekistan (n=3).The red branches represent sequences from Libya, the blue branches adjacent to the Libyan cluster represent sequences from Ghana and Cameroon. Bootstrap values are shown on the tree.

Supplementary Figure 2

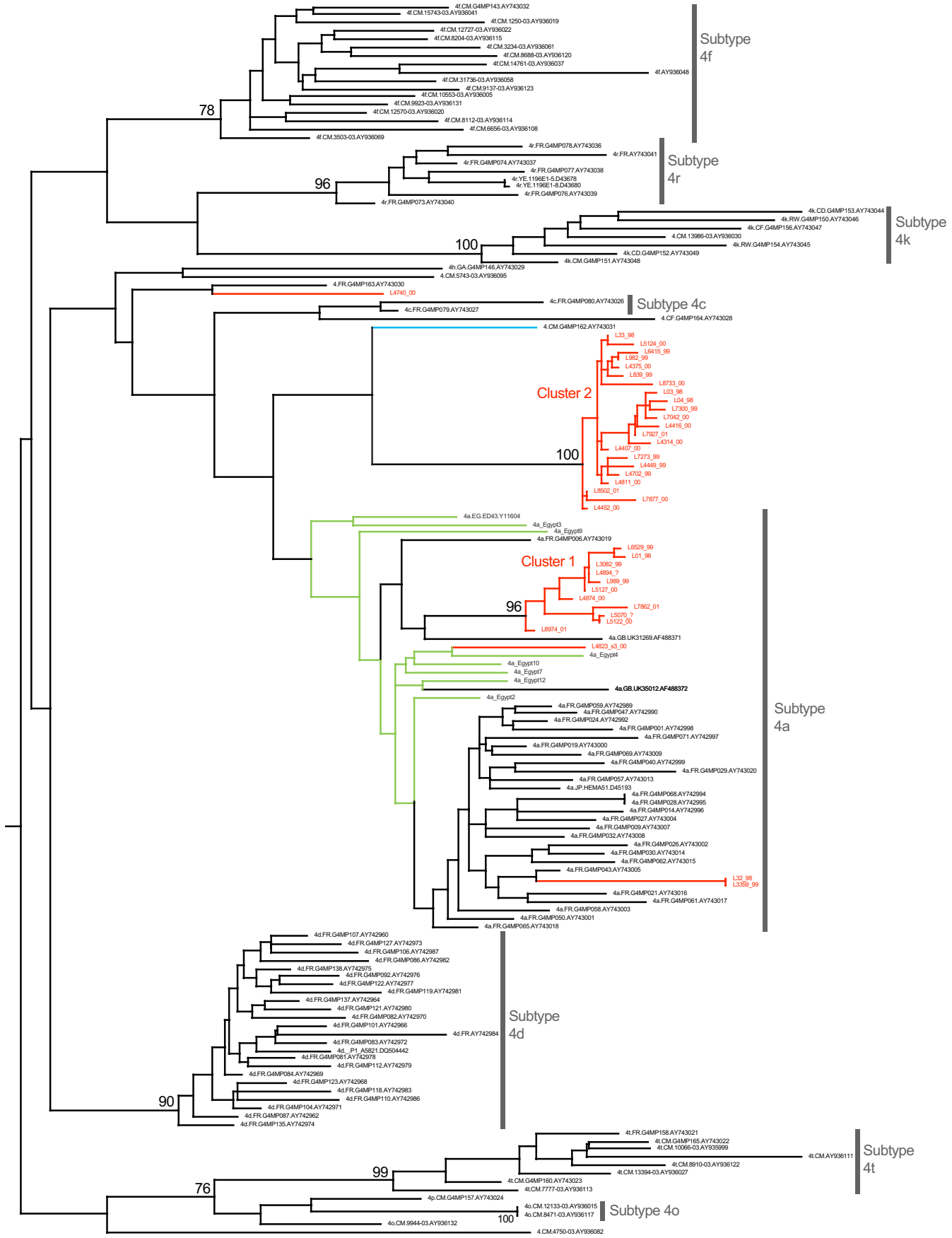
HIV-1 Consensus Majority Rule Bayesian Tree showing posterior probabilities for internal branches of the tree. Sequences from Libya are shown in red; from Ghana and Cameroon in blue.

Supplementary Figure 3



HCV genotype 1 ML phylogenetic tree. Bootstrap values are shown on internal branches. Libyan sequences belonging to cluster 3 are colored red. The grey vertical bars denote HCV subtype classifications.

Supplementary Figure 4



HCV genotype 4 ML phylogenetic tree. Bootstrap values are shown on internal branches. Libyan sequences are coloured red, Egyptian strains are coloured green. The blue coloured strain is from Cameroon. The grey vertical bars denote HCV subtype classifications.