

SUPPLEMENTARY INFORMATION

Authenticity of the data. Several lines of evidence argue strongly that the sequences recovered from the 1960 sample are genuine and not the result of contamination:

(i) The work was performed in laboratories dedicated to working with ancient/degraded samples, under strict controls for contamination, including physical separation of pre- and post-PCR work and numerous extraction and PCR blanks that were consistently negative;

(ii) The results were reproducible across multiple extractions followed by RT-PCR, cloning, and sequencing;

(iii) The results were independently replicated in SMW's laboratory (Northwestern University) using three blinded sub-samples of lymph node tissue; the correct sample was identified as the positive by RT-PCR and verified by cloning and sequencing (**Table S1**);

(iv) 12 short RNA fragments, covering the *gag*, *pol*, and *env* genes, were successfully amplified by RT-PCR. No HIV-1 amplicons larger than 126 nucleotides (including the longer sets used by Zhu *et al.* to amplify ZR59) were recoverable from the sample, as predicted for a Bouin's-fixed paraffin-embedded specimen;

(v) Similarly, the qPCR assay using a human *B2M* RNA marker confirmed that amplifiable RNA was present in the sample but that its quality was relatively low, as expected for paraffin-embedded tissues;

(vi) As observed with ZR59, the relatively short branch length of the HIV-1 sequences from 1960 is consistent with it being an early strain;

(vii) The sample originated from a patient (an adult with apparent lymph node abnormalities) and a tissue type (lymph node) where HIV-1 RNA could plausibly be present and detectable; and

(viii) All of the paraffin-embedded tissues yielded sequences with reasonable topological and branch length properties (*i.e.*, subtype B sequences from the Canadian sample, and subtype D and A sequences from the three linked to the Democratic Republic of the Congo, each closely related to previously published sequences of Congolese origin).

Fig. S1. The 50% majority-rule consensus tree summarizing the Bayesian MCMC phylogenetic analysis conducted using MrBayes v3.1.2. Branch lengths (substitutions/site) represent the average for that branch in tree sample. The posterior probability for each node is indicated. DRC60 (red), ZR59 (black), and the three control sequences recovered from paraffin-embedded specimens (gray) are depicted with bold branches. DRC60 and the two positive controls from the DRC form monophyletic clades with other DRC sequences, while the Canadian control falls as expected within subtype B. Each control sequence has a reasonably long terminal branch length consistent with its date of sampling. The subtype A/A1 ancestral node used for the regression line dating analysis is indicated. The country of sampling and additional details about the reference strains used can be found at <http://hiv.lanl.gov>.

Fig. S2. The maximum clade credibility tree from the relaxed clock Bayesian MCMC analysis conducted using BEAST v1.4.7. This is the same tree as shown in **Fig. 2**, but in rectangular form and with sequence names indicated. The country of sampling and additional details about the reference strains used can be found at <http://hiv.lanl.gov>.

Fig. S3. Bayesian skyline plot of HIV-1 group M. The plot begins at the median posterior TMRCA (1908). The bold line traces the inferred median effective population size over time with the 95% HPD shaded in blue.

Table S1. Primers and sequence-confirmed amplification results.

Primer	Primer Sequence 5'-3'	Frag- ment #	DRC 60	BE81	BE85	CAN97
HIVG1 (F) ^a HIVG2 (R)	ACCCACCTATCCCAGTAGGAGAAAT GGTCCTTGCTTATGTCCAGAATGC	1	(+) ^c	(+)	(+)	(+)
Pol3290F ^b Pol3415R	GCCAGAAAAAGACAGCTGGACTGTCAA AGAGCTTTGGYTCCCCTAAGG	2	(+) ^c	(+)	(+)	(+)
env6371F env6442R	CACCACTCTATTTTGTGCATCAG GCATGTGTAGCCCAGACATTAT	3	(+)	(+)	(+)	(-)
env6445F env6542R	GTGTACCCACAGACCCCAAC CTCATGCATTTGTTCTACCATGT	4	(+)	(+)	(+)	(-)
env6828F env6890R	ACACAGGCTTGTCCAAAGGT ACCAGCTGGGGCACAATAAT	5	(+) ^c	(-)	(-)	(-)
env7468F env7535R	CACTCCCATGCAGAATAAAACA AGGGGCATACATTGCTTGTC	6	(+) ^c	(+)	(+)	(-)
env7717F ^d env7796R	CCACCAAGGCAAAGAGAAGA TCCCAAGAACCCAAGGAAC	7	(-)	(+)	(-)	(+)
env7899F env7977R	ATAGAGGCGCAACAGCATCT TTTCCACAGCCAGGACTCTT	8	(-)	(-)	(-)	(-)
env8047F env8112R	TGCCCTGGAACCTAGTTGG CCATCCAGGTCATGTTCTCC	9	(+) ^c	(+)	(+)	(-)
env8423F env8498R	CGAAGAAGAAGGTGGAGAGC GTCCCAGGCAAGTGCTAAGA	10	(+) ^c	(+)	(-)	(-)
env7717F ^{d,e} env7805R	CCACCAAGGCAAAGAGAAGA TCCTGCTGCTCCTAAGAACC	11	(+)	N/D ^f	N/D	N/D
env7771F ^{d,e} env7846R	GAGCTGTCTTCCTTGGGTTCT GCCTGTACCGTCAGCGTTA	12	(+)	N/D	N/D	N/D
env7835F ^{d,e} env7937R	GACGGTACAGGCCAGACAAT CCAGACCGTGAGTTTCAACA	13	(+)	N/D	N/D	N/D
env7890F ^{d,e} env7960R	CTGAGGGCTATAGAGGCTCAAC CTTGCCTGGAGCTGTTTAATG	14	(+) ^c	N/D	N/D	N/D

^aPreviously published [Boni, J. & Schupbach, J. *J. Virol. Methods* **42**, 309 (1993)].^bThe remaining primer names correspond to the primer's 5' base location in the HXB2 reference strain.^cIndependently replicated in SMW's laboratory, Northwestern University.^dOverlaps ZR59 sequence.^eDesigned specifically for HIV-1 group M subtype A (after DRC60 was determined to be A-like).^fN/D, no data: amplification not attempted.

Fig. S1

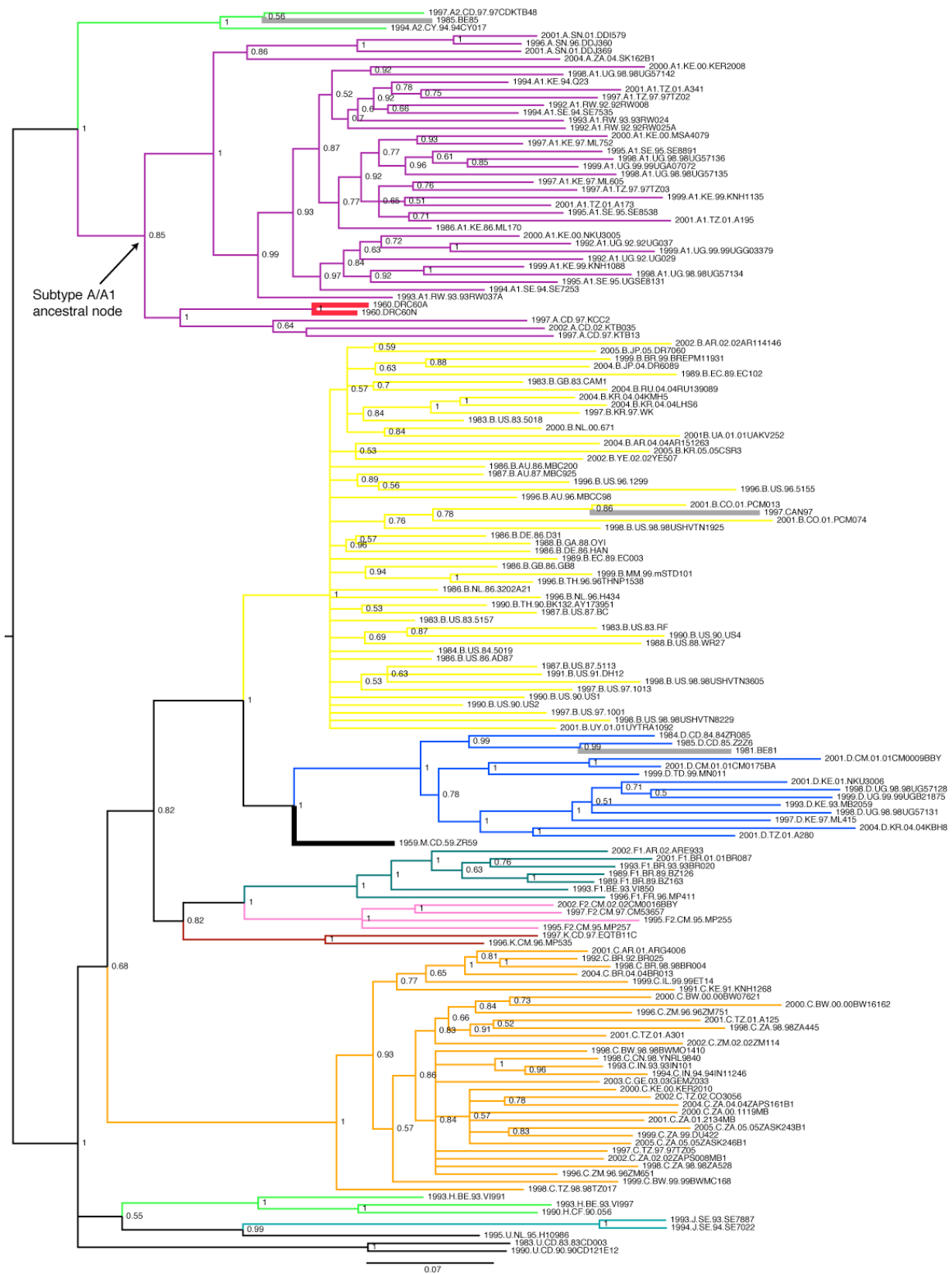


Fig. S2



Fig. S3

