

SUPPLEMENTARY TABLES

Supplementary Table S1. HIV/AIDS estimates in Central America, 2010

	Guatemala	Belize	El Salvador	Honduras	Nicaragua	Costa Rica	Panama
Total Population	13.6 million	315,000	6.1 million	7.9 million	6.0 million	4.5 million	3.4 million
Number of people living with HIV/AIDS	62,000	6,000	34,000	39,000	6,900	9,800	20,000
Number of HIV/ AIDS cases in 2009	1,881	455	1,891	1,060	866	384	1,522
Adult HIV prevalence	0.8%	2.3%	0.8%	0.8%	0.2%	0.3%	0.9%
FSW HIV prevalence	3-7.8%	not known	5.7%	5.5%	1.9%	0.11%	2%
MSM HIV prevalence	11.5-18.3%	not known	10.8%	9.7%	7.9%	12.7%	10.6%
ART coverage rate in 2009	63%	57%	66%	47%	62%	86%	53%

Sources: USAID Report: HIV/AIDS profile for the Central American Region, February, 2011. UNGASS Report, Guatemala, 2010. TB, HIV/AIDS and other STs Programme Report, Belize, 2010. MSPAS, El Salvador, 2009. HIV/AIDS National Report, Honduras, June 2010. ECVC, Honduras, 2006. ECVC, Nicaragua, 2009. HIV/AIDS National Report, Costa Rica, 2009. Epidemiological Surveillance MINSAs, MSPAS, Panama, 2010.

Supplementary Table S2. Clustering results from executing phyloPart on the maximum-likelihood tree containing Central America + control sequences, considering different percentile thresholds and selecting clusters with a reliability >90% (by Shimodaira-Hasegawa test).

Threshold percentile	0.1	1st	5th	10th	15th	25th
No. of clusters	84	108	141	143	135	105
Overall median	2 (2-3)	2 (2-3)	2 (2-3)	2 (2-3)	2 (2-5)	3 (2-5)
No. of sequences/ per/cluster						
Absolute threshold	0.02573	0.04085	0.05496	0.06248	0.06775	0.07558
(nucleotide substitution/site)						

Supplementary Table S3. Total number of Blast searched HIV-1 *pol* sequences included in the Central American phylogenetic analysis.

Country	No. of Blast sequences included in the analysis	Total no. of control sequences included in the analysis	Total no. of HIV-1 <i>pol</i> sequences in LANL
Central American neighboring countries			
US	100	317	15,318
Canada	4	4	1,914
Mexico	0	16	57
Bahamas	0	13	14
Cuba	0	6	211
Dominican Republic	0	3	157
Haiti	0	15	16
Jamaica	1	9	106
Trinidad & Tobago	0	5	30
Argentina	0	4	3,480
Brazil	4	30	4,356

Colombia	0	2	97
Uruguay	0	1	26
Venezuela	6	10	489

Non-Central

American

neighboring countries

Albania	1	1	24
Australia	1	16	353
Azerbaijan	0	1	3
Belgium	0	6	255
China	1	13	1,360
Cyprus	1	7	89
Czech Republic	0	36	765
Denmark	0	11	344
Georgia	0	2	11
Germany	0	28	983
Greece	0	5	86
France	7	45	911

Ireland	0	1	6
Italy	2	48	2,805
Japan	0	45	499
Latvia	1	5	20
Netherlands	0	5	111
Portugal	1	7	297
Russia	0	1	95
Taiwan	0	1	355
Thailand	0	3	47
South Korea	1	8	328
Spain	5	45	1,055
Sweden	0	6	165
Switzerland	0	3	486
Ukraine	0	1	55
United Kingdom	1	19	1,130
Yemen	0	3	9
Total	137	807*	38,918

Note: HIV-1 *pol* control sequences were obtained by BLAST searches in Los Alamos National Laboratory (LANL) in which the closest match to each of our Central American sequences were

retrieved. The analyses also included other sequences from Central American neighboring countries (Latin American and the Caribbean countries) and the rest of the world.

*This total number includes the 67 Central American neighboring countries, including all available Haitian sequences and 740 sequences from the rest of the world.

Supplementary Table S4. Bayes factor between strict (null hypothesis) and relaxed molecular clock (alternative hypothesis).

Model	Clock ^a	Marginal likelihood	BF ^b
Constant	SC	- 24322.20	
	RC	- 24234.18	176.04
BSP	SC	-24315.539	
	RC	-24245.138	140.80

a. The selected molecular clock model (H_0 = Null hypothesis, H_1 = Alternative hypothesis) is highlighted in gray.

b. BF=Bayes Factor. $6 > BF > 2$ indicates positive evidence against the null hypothesis; $10 > BF > 6$ indicates strong evidence against the null hypothesis, $BF > 10$ indicates very strong evidence against the null model.

Supplementary Table S5. Likelihood mapping analysis of HIV-1 *gag*, *pol* and *env* data sets.

Study	Data set *	Star-like signal **
	<i>Env</i>	5.3
	<i>env1</i>	5.7
	<i>env2</i>	3.4
Gilbert et al. (11)	<i>env3</i>	5.6
	<i>env4</i>	5.6
	<i>env5</i>	6.1
	<i>gag</i>	16.6
	<i>gag1</i>	18.8
<hr/>		
	<i>Full data set</i>	52.2
	<i>Clade I</i>	41
Present study	<i>Clade II</i>	26.9
	<i>Clade III</i>	19.5
	<i>Clade IV</i>	25.1
	<i>Clade V</i>	31.9

* *Env* and *gag* data sets are the ones used in the Gilbert et al. (11) paper. Full data set refers to the *pol* alignment analyzed in the present paper including 625 Central American

sequences and 740 randomly selected reference sequences. Clades I-V refer to five sub-alignments, each one including only sequences from one of the supported Central American clades shown in Figure 1.

** Star-like signal is given as percentage of dots in the central area of the likelihood map (37) and represents the amount of phylogenetic noise in the data.

SUPPLEMENTARY FIGURES

Supplementary Figure S1. Maximum likelihood phylogenetic analysis of HIV-1 group M subtype B *pol* sequences. Phylogenetic tree showing each strain labelled as follows: *subtype.country.year.accession number[or strainID for newly sampled strains]*. The tree is the same as the one given in Figure 1 and branches were colored according to the same color legend given in Figure 1.



Supplementary Figure S2. Phylogenetic trees of HIV-1 group M subtype B *pol* sequences using a sub-sample of the reference sequences. Bayesian maximum clade credibility trees were inferred for five datasets, each one including the 21 available Haitian sequences and a random sub-sample of sequences from North America (42 strains), Central America (21 strains), South America (42 strains) and Europe/Oceania (42 strains). The numbers along the monophyletic branches correspond to approximate likelihood ratio test (aLRT) SH-like values. Branch lengths in nucleotide substitutions per site were scaled according to the bar at the bottom of each tree. The first tree is the same as the one shown in Figure 2.









