Supplementary Table S1. Multi-class drug-resistance-associated mutations detected among all cases (n=3904)

Inhibitor	NRTI-R	NNRTI-R	PI-R	n	Year collected
NRTI+NNRTI	M184V	K103N		1	2011
	M41L L74I L210W	K103N		1	2011
NRTI+PI	M41L D67N T215C		V32I M46I V47V L90M	1	2007
	M41L D67N M184V L210W T215S		V32I M46IL I53L L54V V82A L90M	1	2008
	D67G M184V		N88S	1	2008
	K65R		M46I	1	2010
	M41L		N88D	1	2010
	M184V		D30N M46I V82A N88D	1	2012
	V77L		M46I	1	2012
	T215S K219Q		M46I	1	2012
NNRTI+PI		K103N	M46L	3	2010
		K103N	M46L	1	2012
		K103N	M46IL	1	2012
		V106M	M46I	1	2012
3-CLASS	M41L T215D	Y181C	M46I I84V	1	2009
	K219Q	K103N Y181C	I54V	1	2012

NRTI, nucleotide reverse transcriptase inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; PI, protease inhibitor

Supplementary Table S2. Drug-resistance-associated mutations detected each year among recent and long-term seroconverters

A) Recent seroconverters (n=468)

		2007		**2008**		**2009**		**2010**		**2011**		**2012**	
	n=	68		99		64		76		82		79	
NRTI	M41L									2	(2.4%)	1	(1.3%)
	D67E			1	(1.0%)								
	T69D							1	(1.3%)				
	L210W									2	(2.4%)		
	T215X	4	(5.9%)	1	(1.0%)	4	(6.3%)	2	(2.6%)	4	(4.9%)	3	(3.8%)
	K219QR					1	(1.6%)					1	(1.3%)
NNRTI	K103N									1	(1.2%)		
	P225H									1	(1.2%)		
PI	D30N									1	(1.2%)		
	M46IL	1	(1.5%)	5	(5.1%)	2	(3.1%)	2	(2.6%)	2	(2.4%)	3	(3.8%)
	N88D									1	(1.2%)		

B) Long-term seroconverters (n=935)

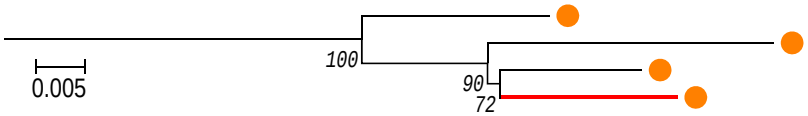
		2007		2008		2009		2010		2011		2012	
	n=	100		142		129		166		189		209	
NRTI	M41L	1	(1.0%)					1	(0.6%)	1	(0.5%)		
	K65R							1	(0.6%)				
	D67N	1	(1.0%)									1	(0.5%)
	K70R							1	(0.6%)				
	F77L					1	(0.8%)						
	M184V									1	(0.5%)	1	(0.5%)
	L210W	1	(1.0%)							1	(0.5%)		
	T215X	4	(4.0%)	2	(1.4%)	4	(3.1%)	5	(3.0%)	9	(4.8%)	4	(1.9%)
	K219N											1	(0.5%)
NNRTI	K101E							1	(0.6%)	1	(0.5%)	1	(0.5%)
	K103NS	2	(2.0%)	2	(1.4%)	1	(0.8%)	5	(3.0%)	2	(1.1%)	3	(1.4%)
	V106M											1	(0.5%)
	G190A											1	(0.5%)
	P225H							1	(0.6%)				
	M230L											1	(0.5%)
PI	D30N							4	(2.4%)				
	M46IL	1	(1.0%)	5	(3.5%)	4	(3.1%)	6	(3.6%)	1	(0.5%)	12	(5.7%)
	I85V					1	(0.8%)						
	N88D							3	(1.8%)				
	L90M					1	(0.8%)						

NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; PI, protease inhibitor

FIGURE S1. Subtrees of a maximum likelihood tree in which sequences with drug-resistant mutations are concentrated.

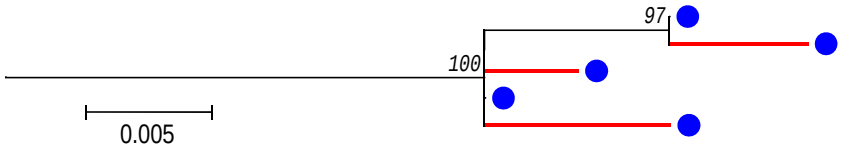
Subtrees comprising of sequences with A) D30N/N88D, B) T215 revertant, and C) K103N are zoomed up from the maximum likelihood tree (Fig. 2). Recent seroconverters are shown in red branches and long-term seroconverters in black. Number left of a node is the bootstrap values >70.

A) D30N/N88D

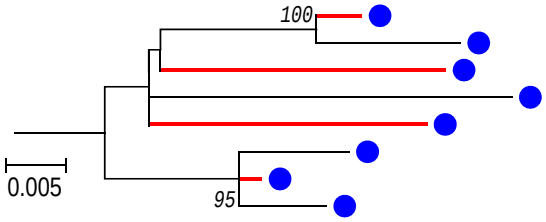


B) T215 revertant

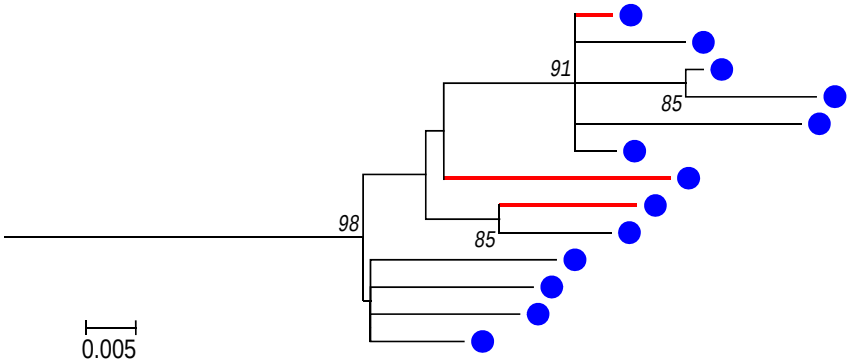
T215E



T215D/E



T215D/E



C) K103N

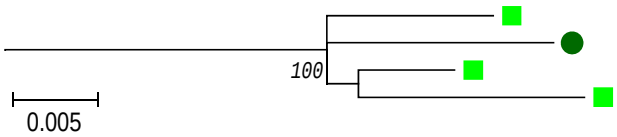
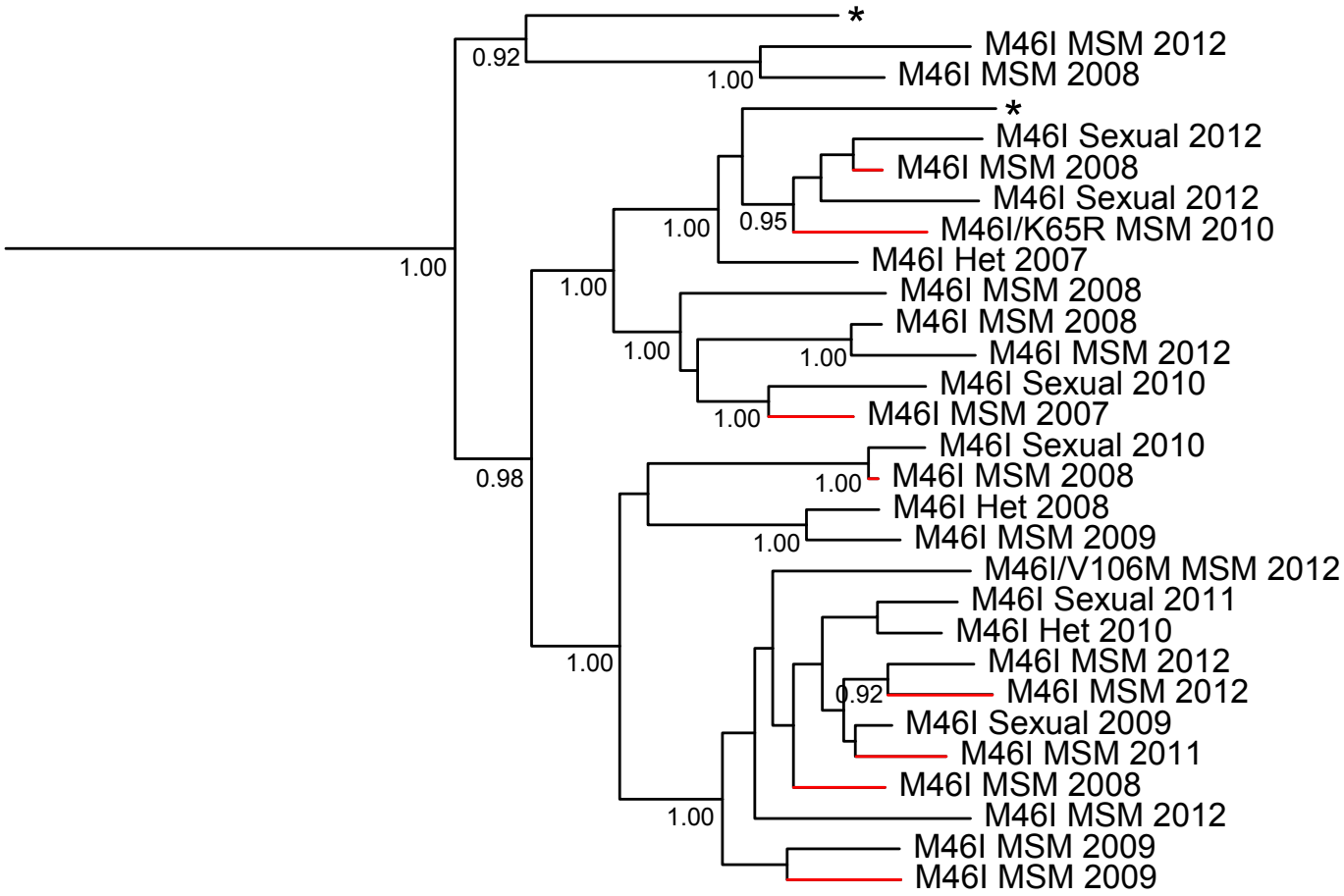


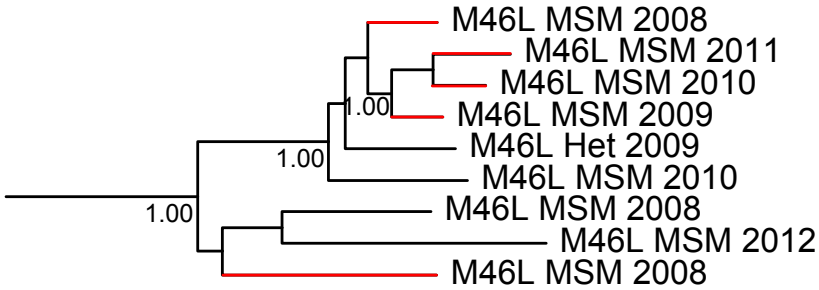
FIGURE S2. Subtrees of a Bayesian maximum clade credibility tree in which sequences with drug resistant mutations are concentrated.

Subtrees comprising of sequences with A) M46I, B) M46L, C) D30N/N88D, D) T215 revertant and E) K103N. Number left of a node is the posterior probability (>0.95) of corresponding cluster. Detected drug-resistant mutation, risk behavior, and the year of sample collection for each sample are shown. Recent seroconverters are indicated with red branches and long-term seroconverters in black. Asterisks depict sequences in which the M46I mutation was not detected. Het: heterosexual risk behavior; MSM: men who have sex with men; Sexual: non-specified sexual risk behavior.

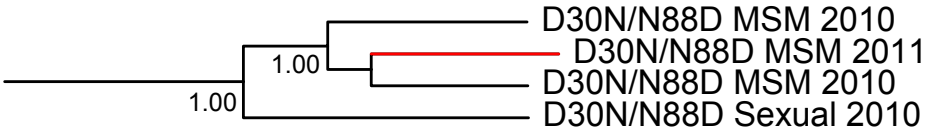
A) M46I



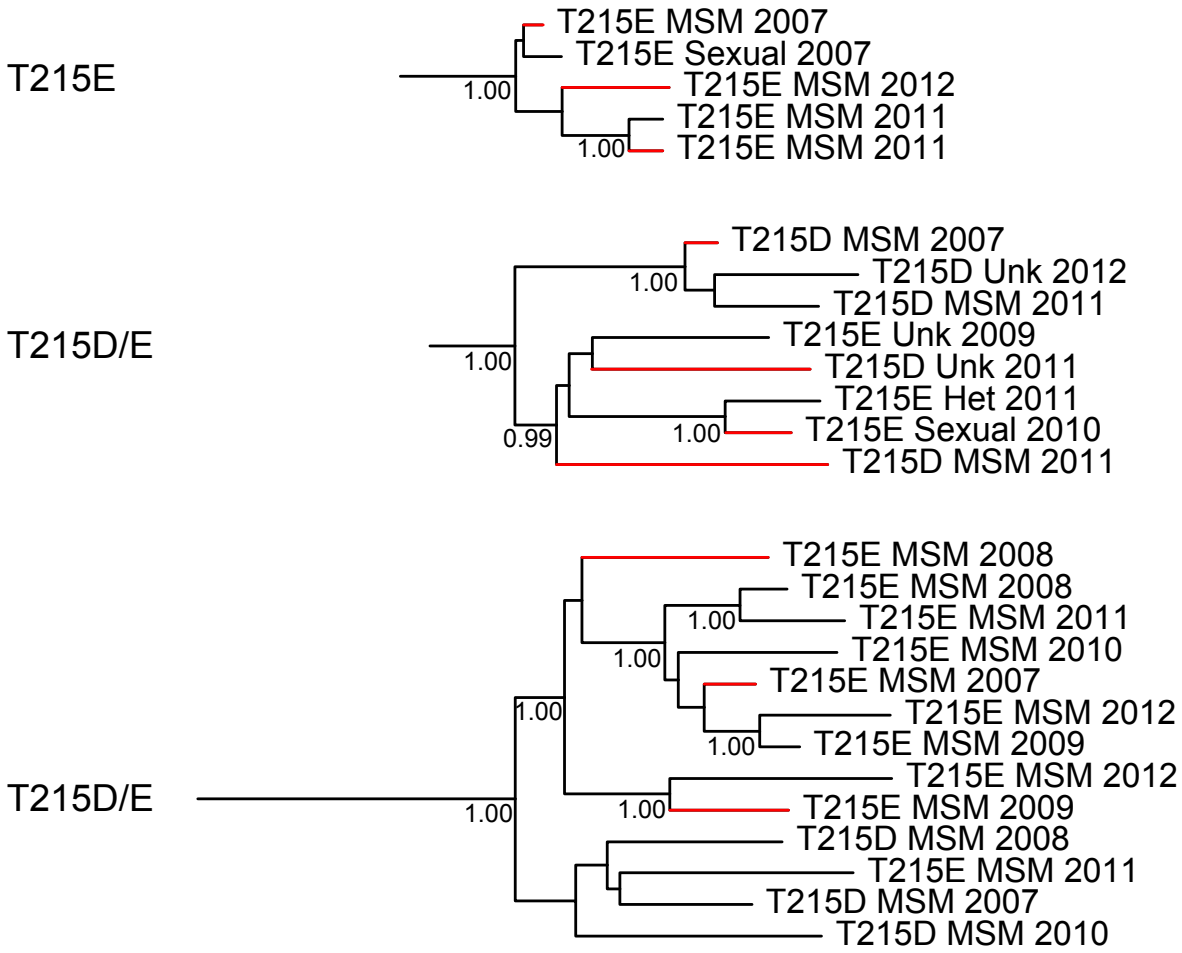
B) M46L



C) D30N/N88D



D) T215 revertant



E) K103N