Supplementary Materials:

Supplemental Digital Content 1. Phylogenetic tree analysis of the full-length newly characterized MSM *env* sequences. pdf

Supplemental Digital Content 2. Phylogenetic relationships between the Beijing MSM sequences and other partial *env* sequences in China. pdf

Supplemental Digital Content 3. Phylogenetic relationships between the MSM and non-MSM *env* sequences in the Beijing city. pdf

Supplemental Digital Content 4. Determination of autologous and heterologous neutralization to multiple T/F Env pesudoviruses in subjects BJOX25 and BJOX28. pdf

Supplemental Digital Content 5. Evolutionary characteristics of the MSM *env* sequences in Beijing. Pdf

Supplemental Digital Content 6. Detection of autologous neutralizing antibodies in longitudinal plasma samples. pdf

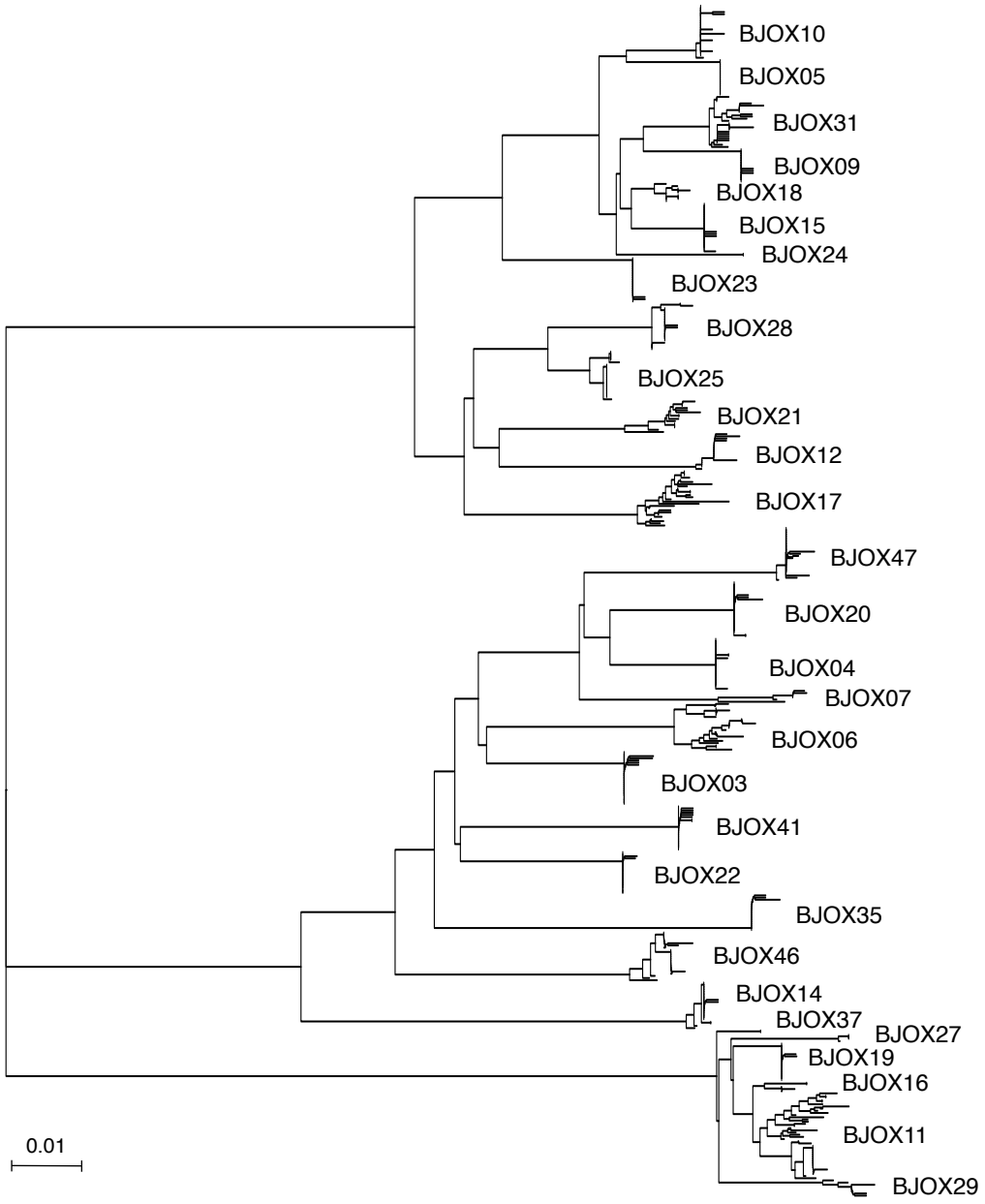


Fig. S1. Phylogenetic tree analysis of the full-length newly characterized MSM *env* sequences. The phylogenetic tree was constructed with the *env* SGA sequences from 30 HIV-1 infected MSM individuals using the neighbor-joining method and the Kimura two-parameter model. The scale bar represents 0.01 nucleotide substitutions per site. Asterisks indicate bootstrap values in which the cluster to the right is supported in 80% or more replicates (out of 1000). All sequences from each individual clustered tightly together.

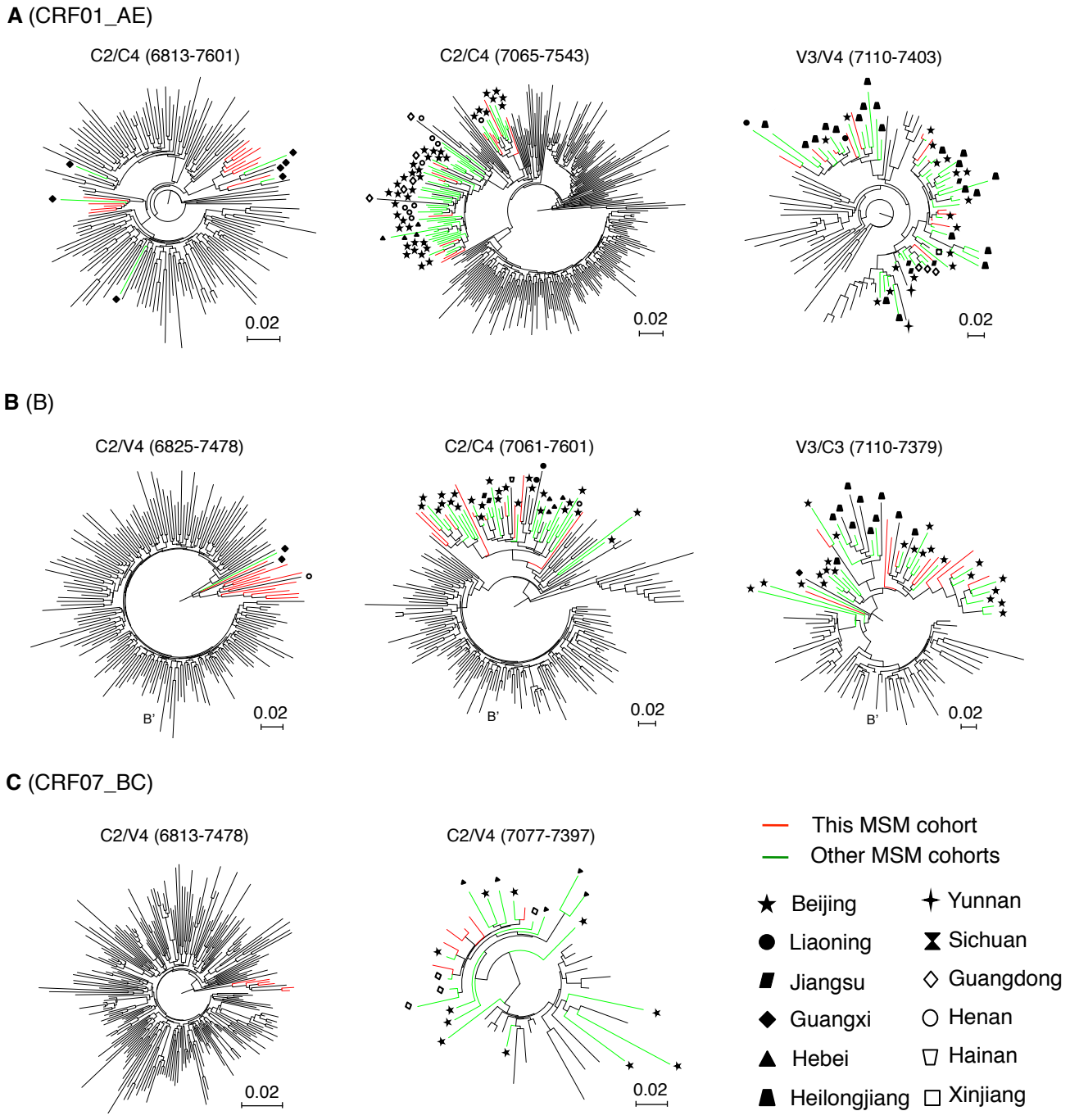


Fig. S2. Phylogenetic relationships between the Beijing MSM sequences and other partial *env* sequences in China. Partial gp120 *env* sequences (>300bp) in China available from Los Alamos HIV-1 Sequencer Database were analyzed together with the newly obtained sequences from the Beijing MSM cohort for CRF01_AE (A), subtype B (B), CRF07_BC (C). Since those sequences did not overlap each other, various neighbor-joining trees were constructed by including as many sequences as possible and maximally preserving the sequence length for each sequence fragments. The final consensus overlapping region used for phylogenetic tree analysis of each region is indicated in brackets. The MSM sequences from this Beijing cohort and other cohorts are indicated by red and green, respectively. Non-MSM sequences are shown in black. The MSM sequences from different provinces are indicated with different symbols.

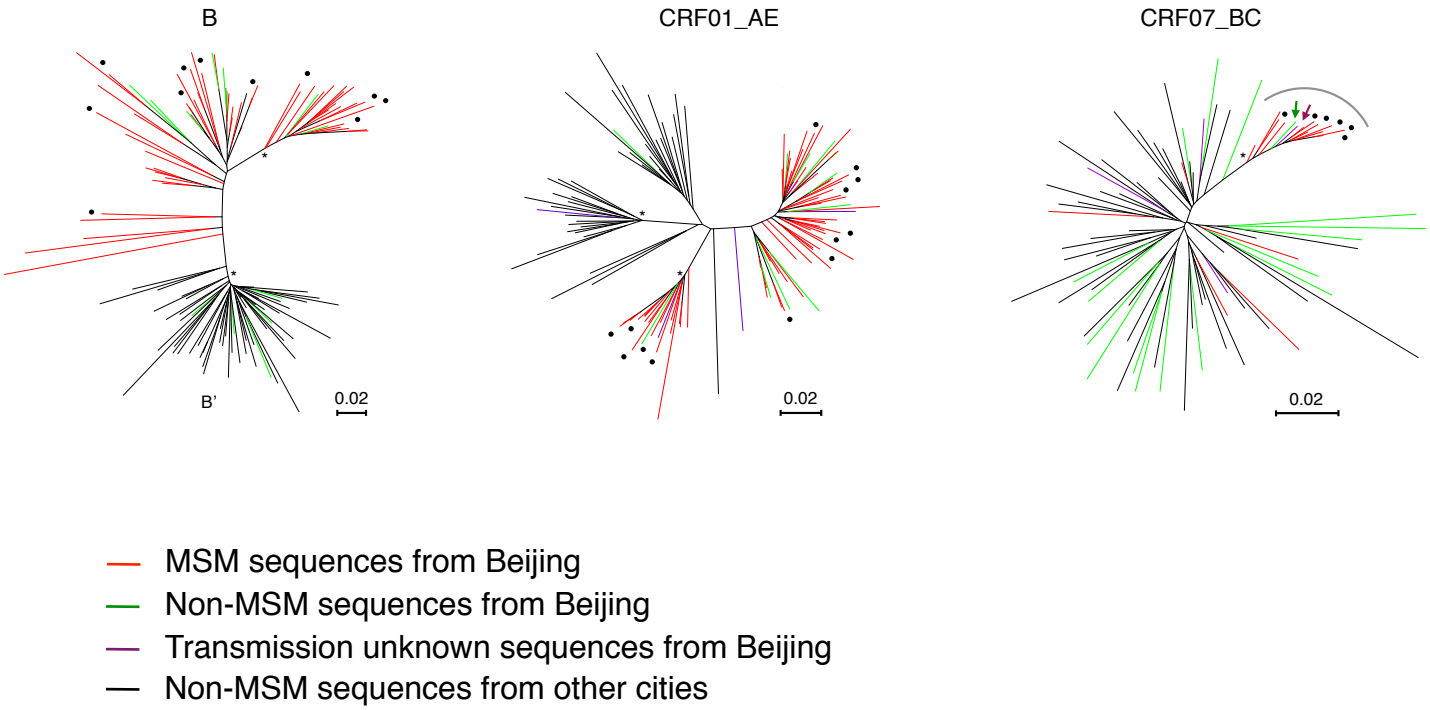
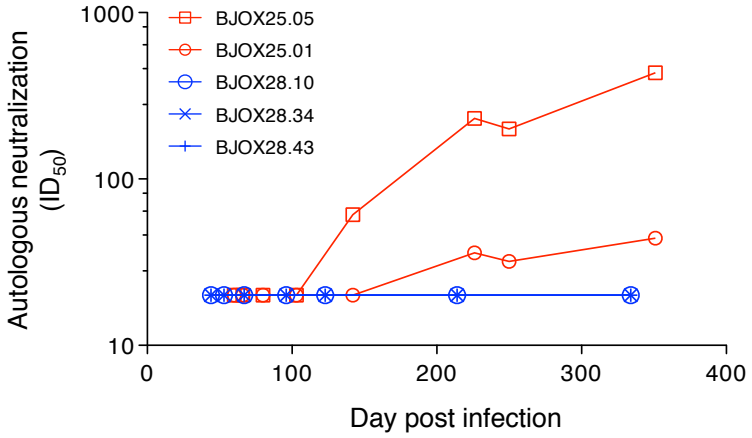


Fig. S3. Phylogenetic relationships between the MSM and non-MSM *env* sequences in Beijing. The phylogenetic relationship of the additional partial *env* sequences (324-1128 bp) obtained from non-MSM individuals in Beijing and the MSM *env* sequences from the same city were determined. The phylogenetic trees were constructed using the neighbor-joining method and the Kimura two-parameter model. The scale bar represents 0.02 nucleotide substitutions per site. Asterisks indicate bootstrap values in which the node is supported in 80% or more replicates (out of 1,000). The consensus C2C4 overlapping region (nt7077-nt7512) was used for phylogenetic tree analysis. Sequences from MSM and non-MSM in Beijing are indicted by red and green, respectively. Sequences from this Beijing MSM cohort are indicated by black dots. The sequences with unknown transmission routes are indicated by purple and representative non-MSM sequences are shown in black. One non-MSM sequence and one sequence with unknown transmission route that clustered within a tight Beijing MSM sequence clade (CRF07_BC) are indicated by green and purple arrows, respectively.

A



B

	CRF01_AE										B											CRF07_BC					
	BJOX05 (d504)	BJOX09 (d495)	BJOX10 (d453)	BJOX18 (d393)	BJOX23 (d422)	BJOX31 (d365)	BJOX12 (d535)	BJOX21 (d530)	BJOX25 (d351)	BJOX28 (d334)	BJOX03 (d533)	BJOX04 (d649)	BJOX06 (d542)	BJOX07 (d549)	BJOX14 (d442)	BJOX20 (d378)	BJOX22 (d375)	BJOX35 (d85)	BJOX41 (d359)	BJOX46 (d350)	BJOX47 (d273)	BJOX11 (d241)	BJOX16 (d274)	BJOX19 (d387	BJOX27 (d315)	BJOX29 (d402)	BJOX37 (d380)
BJOX25.01	<20	nt	<20	<20	<20	<20	<20	53	135	<20	nt	nt	<20	<20	<20	<20	nt	nt	26	<20	nt	nt	37	nt	nt	<20	nt
BJOX25.05	<20	<20	<20	<20	<20	42	<20	38	506	<20	<20	<20	70	<20	<20	<20	<20	<20	37	<20	<20	<20	67	<20	<20	<20	75
BJOX28.10	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20
BJOX28.34	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	39	<20	<20	<20	<20	<20	<20	<20	<20
BJOX28.43	<20	<20	<20	<20	<20	<20	<20	<20	131	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20	<20

Days post infection are indicated in brackets.

Figure S4. Determination of autologous and heterologous neutralization to multiple T/F Env pesudoviruses in subjects BJOX25 and BJOX28. (A) Neutralizing antibody titers in plasma were determined by measuring the luciferase activity in TZM-bl cells. Two BJOX25 T/F Env pseudoviruses (red) and three BJOX28 T/F Env pseudoviruses (blue) were assayed against autologous longitudinal plasma. (B) Two BJOX25 T/F Env pseudoviruses and three BJOX28 T/F Env pseudoviruses were assayed against heterologous plasma as shown in Fig.4. Reciprocal plasma dilutions at which luciferase activity was reduced by 50% (ID_{50}) relative to no plasma control wells are shown: Red: >1: 500; Orange: 1: 100-1: 500; Green: 1: 20-<1: 100; White: <1: 20 or nt (not tested)

Table S1. Evolutionary characteristics of the MSM *env* sequences in Beijing

Genotype	Cluster	Value of the evolutionary parameter[a]							
		HKY+Γ_4 constant size				GTR+Γ_4 constant size			
		Relaxed Model		Strict Model		Relaxed Model		Strict Model	
		Evolutionary rate[b]	tMRCA[c]	Evolutionary rate[b]	tMRCA[c]	Evolutionary rate[b]	tMRCA[c]	Evolutionary rate[b]	tMRCA[c]
Subtype B	I	6.6 (4.4, 8.9)	1985 (1975-1993)	5.7 (4.7, 6.8)	1982 (1977-1988)	6.2 (3.9, 8.5)	1984 (1973-1994)	6.0 (5.0, 7.0)	1984 (1979-1988)
CRF01_AE	I	7.3 (5.3, 9.3)	1994 (1989-1998)	6.4 (5.3, 7.5)	1992 (1989-1995)	7.2 (5.3, 9.0)	1994 (1990-1998)	6.2 (5.2, 7.3)	1992 (1989-1995)
	II		1996 (1992-1999)		1995 (1992-1997)		1996 (1992-1999)		1995 (1992-1997)
CRF07_BC	I	7.7 (6.0, 9.3)	2002 (2000-2004)	7.2 (6.2, 8.2)	2002 (2001-2003)	7.5 (6.3, 9.9)	2003 (2001-2004)	7.0 (6.0, 7.9)	2002 (2001-2003)

[a] The 95% highest posterior density (HPD) confidence intervals are given in parentheses.

[b] Estimated evolutionary rates ($\times 10^{-3}$ nucleotide substitutions/site/year).

[c] Time to the most recent common ancestor (tMRCA: year).

Table S2. Detection of autologous neutralizing antibodies in longitudinal plasma samples

Virus	DPI	ID_{50}
B_BJOX03.23	238	<20
	323	<20
	407	<20
	533	125
01_BJOX05.11	42	<20
	49	<20
	41	<20
	77	<20
	106	<20
	190	114
	303	460
	387	548
	504	551
B_BJOX06.13*	125	<20
	132	<20
	146	<20
	175	<20
	194	<20
	276	35
	367	46
	542	62
B_BJOX07.02	113	<20
	120	<20
	134	<20
	155	<20
	186	<20
	269	265
	353	163
	437	99
	549	235
01_BJOX09.02	59	<20
	102	<20
	265	156
	495	288
01_BJOX10.07	53	<20
	60	<20
	74	<20
	106	<20
	137	<20
	232	101
	319	240
	453	377
07_BJOX11.01*	108	<20
	122	<20
	136	<20
	170	<20
	189	47
	241	122
01_BJOX12.04	108	<20
	114	<20
	128	<20
	156	<20
	183	<20
	282	59
	367	77
	449	82
	535	281
B_BJOX14.09	162	<20
	176	39
	232	2122
	351	2849
	379	3002
	442	3915
01_BJOX15.11	68	<20
	105	<20
	115	<20
01_BJOX18.02*	54	<20
	68	<20
	74	<20
	108	<20
	138	<20
	227	23
	286	76
	393	177
07_BJOX19.02	44	<20
	79	<20
	93	<20
	107	<20
	254	<20
	387	<20
B_BJOX20.03	42	<20
	49	<20
	63	<20
	91	<20
	114	<20
	202	95
	279	414
	378	508
B_BJOX22.02	49	<20
	59	<20
	74	<20
	103	<20
	136	<20
	199	131
	267	<20
	375	<20
01_BJOX23.05	51	<20
	58	<20
	72	<20
	135	64
	212	223
	422	681
01_BJOX25.01	60	<20
	66	<20
	80	<20
	103	<20
	142	20
	226	36
	250	32
	351	44
01_BJOX25.05	60	<20
	66	<20
	80	<20
	103	<20
	142	61
	226	231
	250	200
	351	435
07_BJOX27.02*	162	<20
	174	<20
	203	<20
	236	<20
	315	<20
01_BJOX28.10	44	<20
	53	<20
	67	<20
	97	<20
	123	<20
	214	<20
	334	<20
01_BJOX28.34	44	<20
	53	<20
	67	<20
	97	<20
	123	<20
	214	<20
	334	<20
01_BJOX28.43	44	<20
	53	<20
	67	<20
	97	<20
	123	<20
	214	<20
	334	<20
07_BJOX29.03*	110	<20
	117	<20
	130	<20
	162	<20
	191	<20
	275	105
	402	1798
01_BJOX31.02*	112	<20
	232	41
	365	456
B_BJOX35.04	37	<20
	43	<20
	53	<20
	85	<20
07_BJOX37.02	163	<20
	171	<20
	184	30
	211	55
	238	56
	270	44
	380	81
B_BJOX41.16*	170	<20
	177	<20
	191	<20
	227	99
	359	568
B_BJOX46.26*	157	<20
	267	<20
	350	141
B_BJOX47.20	158	<20
	165	<20
	184	<20
	212	<20
	273	<20

*Virus for which T/F *env* sequences can not be inferred