Dataset	genomic region	PS	SS	#days to first sampling	#time points recipient	# sequences recipient	# sequences donor
BC							
Andreo *et al.* [1]	env	exp	exp	183	2	13	16
Blanchard *et al.* [2]	env	exp	exp	1149	1	2	6
	gag	con	con	”	”	2	5
Cornelissen *et al.* [3]	env	con	con	21	2	20	8
	gag	exp	exp	”	”	20	9
Diaz-Zhang *et al.* [4]							
source→recipA	env	con	con	791	4	39	61
source→recipB	”	”	”	670	5	56	”
Kao *et al.* [5]	env	log	log	316	1	11	31
Metzker *et al.* [6]	env	exp	exp	420	1	50	50
Wolfs *et al.* [7]	env	con	con	30	1	8	9
	gag	exp	exp	”	1	10	8
			mean	447,5	2,1	19,3	20,3
			median	368	1,5	12	9
HET							
Boeras *et al.* [8]							
RW36	env	exp	exp	5	1	41	59
RW56	env	exp	exp	31	1	39	58
ZM201	env	log	log	14	1	46	93
ZM216	env	log	log	14	1	59	75
ZM221	env	exp	log	14	1	72	71
ZM238	env	exp	exp	47	1	38	72
ZM242	env	exp	exp	14	1	31	44
ZM292	env	exp	exp	14	1	39	71
Derdeyn *et al.* [9]							
13	env	con	con	123	1	16	23
53	env	exp	exp	23	1	19	34
55	env	exp	exp	14	1	16	36
71	env	exp	exp	14	1	11	10
83	env	log	log	20	1	11	12
106	env	con	con	53	1	13	12
109	env	exp	exp	14	1	13	13
135	env	exp	exp	26	1	20	32
Haaland *et al.* [10]							
RW19	env	exp	exp	5	1	40	40
RW35	env	log	log	5	1	40	37
RW41	env	exp	exp	5	1	40	40
RW53	env	exp	exp	14	1	42	36
RW57	env	exp	exp	5	1	42	40
RW66	env	exp	exp	14	1	41	42
ZM190	env	log	log	14	1	48	35
ZM198	env	exp	exp	20	1	41	42
ZM205	env	log	log	16	1	32	45
ZM229	env	exp	exp	5	1	27	22
ZM243	env	exp	exp	14	1	45	36
ZM248	env	log	log	20	1	57	56
Hayman *et al.* [11]							
don1→3	env	con	con	1462	1	3	10
don1→19	”	”	”	367	1	3	”
don1→22	”	”	”	1827	1	6	”

Continued on next page

Table 1 – continued from previous page

Dataset	genomic region	PS	SS	#days to first sampling	#time points recipient	# sequences recipient	# sequences donor
don17→23	env	log	log	913	1	3	6
Liu *et al.* [12]							
D	env	exp	exp	61	1	11	19
E	env	exp	exp	61	1	15	16
F	env	exp	exp	39	1	14	15
G	env	exp	exp	23	1	13	14
I	env	log	log	52	1	9	15
J	env	exp	exp	5	1	10	13
Scaduto *et al.* [13]							
CC 01→02	env	exp	exp	101	1	20	20
01→03	”	”	”	606	1	20	”
01→04	”	”	”	777	1	20	”
01→05	”	”	”	697	1	20	”
01→06	”	”	”	737	1	20	”
01→07	”	”	”	471	1	20	”
CC 01→02	pol	exp	exp	”	”	20	20
01→03	”	”	”	”	”	20	”
01→04	”	”	”	”	”	20	”
01→05	”	”	”	”	”	20	”
01→06	”	”	”	”	”	20	”
01→07	”	”	”	”	”	20	”
WA 04→01	env	exp	exp	132	1	20	20
04→02	”	”	”	161	1	20	”
04→03	”	”	”	511	1	20	”
04→05	”	”	”	205	1	19	”
04→06	”	”	”	207	1	17	”
WA 04→01	pol	exp	exp	”	”	20	20
04→02	”	”	”	”	”	20	”
04→03	”	”	”	”	”	20	”
04→05	”	”	”	”	”	19	”
04→06	”	”	”	”	”	17	”
Wolfs *et al.* [7]	env	exp	exp	5	1	8	8
			mean	203,9	1	24,7	33,7
			median	23	1	20	34
MSM							
Bailey *et al.* [14]	gag	exp	exp	9493	4	38	6
Butler *et al.* [15]							
sourceABC→A	env	exp	exp	83	1	10	53
sourceABC→B	”	”	”	71	1	8	”
sourceABC→C	”	”	”	48	1	5	”
D	env	exp	exp	128	1	13	13
E	env	exp	exp	73	1	17	22
F	env	exp	exp	84	1	19	44
Edo-Matas *et al.* [16]	env	log	log	2994	4	40	23
Edwards *et al.* [17]	env	log	log	242	4	67	33
	gag	log	log	”	4	64	39
Frost *et al.* [18]							
004-007	env	exp	exp	10	1	12	12
206-201	env	exp	exp	10	1	11	12

Continued on next page

Table 1 – continued from previous page

Dataset	genomic region	PS	SS	#days to first sampling	#time points recipient	# sequences recipient	# sequences donor
206-204	env	exp	exp	10	1	12	13
512-558	env	exp	exp	104	1	14	11
512-559	env	exp	exp	104	1	11	11
551-550	env	exp	exp	10	1	12	13
564-557	env	exp	exp	10	1	15	14
Herbeck *et al.* [19]							
1	gag	exp	exp	36	15	304	20
1	env	log	log	”	12	418	21
1	pol	exp	exp	”	12	152	10
2	gag	exp	exp	13	1	18	10
2	env	exp	exp	”	”	18	10
2	pol	exp	exp	”	”	18	10
3	gag	exp	exp	30	1	93	10
3	env	exp	exp	”	”	93	10
3	pol	exp	exp	”	”	93	10
4	gag	exp	exp	20	1	30	38
4	env	exp	exp	”	”	30	38
4	pol	exp	exp	”	”	30	38
Lawson *et al.* [20]	env	exp	exp	5	2	19	10
Li *et al.* [21]							
AD18-AD17	env	exp	exp	10	1	25	17
AD75-AD77	env	con	con	10	1	27	19
AD83-04013240	env	exp	exp	10	1	66	22
LAC-HOBR	env	exp	log	10	1	21	13
Rachinger *et al.* [22]							
pat3→1	env	log	log	153	5	46	65
pat3→2	”	”	”	284	6	61	”
pat3→1	gag	log	log	”	5	46	63
pat3→2	”	”	”	”	7	46	”
pat3→1	pol	exp	exp	”	2	26	24
pat3→2	”	”	”	”	3	15	”
			mean	X	X	X	X
			median	X	X	X	X

Overview of the examined transmission chains. For each article in which we found useful data, we detail from what genomic region the sequences originate (env = Envelope, gag = Gag polyprotein, pol = Polymerase). We also give the number of sequences by recipient and donor and what demographic model came out best by the path sampling (PS) and stepping stone sampling (SS) marginal likelihood estimators (exp = exponential, con = constant, log = logistic). We also also specify the time between the most recent boundary of the transmission interval and the first sampling of the recipient as well as the number of available time points for the recipient. When multiple transmission chains were available, we kept the original naming and specify all details for each chain separately. Similarly, when a chain comprises multiple transmission events, the data are specified per event.

References

[1] Andreo, S.M.S., Barra, L.A.C., Costa, L.J., Sucupira, M.C.A., Souza, I.E.L., Diaz, R.S.: Hiv type 1 transmission by human bite. AIDS Res Hum Retroviruses **20**(4), 349–50 (2004)

[2] Blanchard, A., Ferris, S., Chamaret, S., Guétard, D., Montagnier, L.: Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. J Virol **72**(5), 4537–40 (1998)

[3] Cornelissen, M., Mulder-Kampinga, G., Veenstra, J., Zorgdrager, F., Kuiken, C., Hartman, S., Dekker, J., van der Hoek, L., Sol, C., Coutinho, R.: Syncytium-inducing (si) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/si phenotypically mixed human immunodeficiency virus population. J Virol **69**(3), 1810–8 (1995)

[4] Diaz, R.S., Zhang, L., Busch, M.P., Mosley, J.W., Mayer, A.: Divergence of hiv-1 quasispecies in an epidemiologic cluster. AIDS **11**(4), 415–22 (1997)

[5] Kao, C.-F., Hsia, K.-T., Chang, S.-Y., Chang, F.-Y., Nelson, K., Yang, C.-H., Huang, Y.-F., Fu, T.-Y., Yang, J.-Y.: An uncommon case of hiv-1 transmission due to a knife fight. AIDS Res Hum Retroviruses **27**(2), 115–22 (2011)

[6] Metzker, M.L., Mindell, D.P., Liu, X.-M., Ptak, R.G., Gibbs, R.A., Hillis, D.M.: Molecular evidence of hiv-1 transmission in a criminal case. Proc Natl Acad Sci U S A **99**(22), 14292–7 (2002)

[7] Wolfs, T.F., Zwart, G., Bakker, M., Goudsmit, J.: Hiv-1 genomic rna diversification following sexual and parenteral virus transmission. Virology **189**(1), 103–10 (1992)

[8] Boeras, D.I., Hraber, P.T., Hurlston, M., Evans-Strickfaden, T., Bhattacharya, T., Giorgi, E.E., Mulenga, J., Karita, E., Korber, B.T., Allen, S., Hart, C.E., Derdeyn, C.A., Hunter, E.: Role of donor genital tract hiv-1 diversity in the transmission bottleneck. Proc Natl Acad Sci U S A **108**(46), 1156–63 (2011)

[9] Derdeyn, C.A., Decker, J.M., Bibollet-Ruche, F., Mokili, J.L., Muldoon, M., Denham, S.A., Heil, M.L., Kasolo, F., Musonda, R., Hahn, B.H., Shaw, G.M., Korber, B.T., Allen, S., Hunter, E.: Envelope-constrained neutralization-sensitive hiv-1 after heterosexual transmission. Science **303**(5666), 2019–22 (2004)

[10] Haaland, R.E., Hawkins, P.A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., Manigart, O., Mulenga, J., Keele, B.F., Shaw, G.M., Hahn, B.H., Allen, S.A., Derdeyn, C.A., Hunter, E.: Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype a and c hiv-1. PLoS Pathog **5**(1), 1000274 (2009)

[11] Hayman, A., Moss, T., Simmons, G., Arnold, C., Holmes, E.C., Naylor-Adamson, L., Hawkswell, J., Allen, K., Radford, J., Nguyen-Van-Tam, J., Balfe, P.: Phylogenetic analysis of multiple heterosexual transmission events involving subtype b of hiv type 1. AIDS Res Hum Retroviruses **17**(8), 689–95 (2001)

[12] Liu, Y., Curlin, M.E., Diem, K., Zhao, H., Ghosh, A.K., Zhu, H., Woodward, A.S., Maenza, J., Stevens, C.E., Stekler, J., Collier, A.C., Genowati, I., Deng, W., Zioni, R., Corey, L., Zhu, T., Mullins, J.I.: Env length and n-linked glycosylation following transmission of human immunodeficiency virus type 1 subtype b viruses. Virology **374**(2), 229–33 (2008)

[13] Scaduto, D.I., Brown, J.M., Haaland, W.C., Zwickl, D.J., Hillis, D.M., Metzker, M.L.: Source identification in two criminal cases using phylogenetic analysis of hiv-1 dna sequences. Proc Natl Acad Sci U S A **107**(50), 21242–7 (2010)

[14] Bailey, J.R., O'Connell, K., Yang, H.-C., Han, Y., Xu, J., Jilek, B., Williams, T.M., Ray, S.C., Siliciano, R.F., Blankson, J.N.: Transmission of human immunodeficiency virus type 1 from a patient who developed aids to an elite suppressor. J Virol **82**(15), 7395–410 (2008)

[15] Butler, D.M., Delport, W., Kosakovsky Pond, S.L., Lakdawala, M.K., Cheng, P.M., Little, S.J., Richman, D.D., Smith, D.M.: The origins of sexually transmitted hiv among men who have sex with men. Sci Transl Med **2**(18), 18–1 (2010)

[16] Edo-Matas, D., Rachinger, A., Setiawan, L.C., Boeser-Nunnink, B.D., van 't Wout, A.B., Lemey, P., Schuitemaker, H.: The evolution of human immunodeficiency virus type-1 (hiv-1) envelope molecular properties and coreceptor use at all stages of infection in an hiv-1 donor-recipient pair. Virology **422**(1), 70–80 (2012)

[17] Edwards, C.T.T., Holmes, E.C., Wilson, D.J., Viscidi, R.P., Abrams, E.J., Phillips, R.E., Drummond, A.J.: Population genetic estimation of the loss of genetic diversity during horizontal transmission of hiv-1. BMC Evol Biol **6**, 28 (2006)

[18] Frost, S.D.W., Liu, Y., Pond, S.L.K., Chappey, C., Wrin, T., Petropoulos, C.J., Little, S.J., Richman, D.D.: Characterization of human immunodeficiency virus type 1 (hiv-1) envelope variation and neutralizing antibody responses during transmission of hiv-1 subtype b. J Virol **79**(10), 6523–7 (2005)

[19] Herbeck, J.T., Rolland, M., Liu, Y., McLaughlin, S., McNevin, J., Zhao, H., Wong, K., Stoddard, J.N., Raugi, D., Sorensen, S., Genowati, I., Birditt, B., McKay, A., Diem, K., Maust, B.S., Deng, W., Collier, A.C., Stekler, J.D., McElrath, M.J., Mullins, J.I.: Demographic processes affect hiv-1 evolution in primary infection before the onset of selective processes. J Virol **85**(15), 7523–34 (2011)

[20] Lawson, V.A., Oelrichs, R., Guillon, C., Imrie, A.A., Cooper, D.A., Deacon, N.J., McPhee, D.A.: Adaptive changes after human immunodeficiency virus type 1 transmission. AIDS Res Hum Retroviruses **18**(8), 545–56 (2002)

[21] Li, H., Bar, K.J., Wang, S., Decker, J.M., Chen, Y., Sun, C., Salazar-Gonzalez, J.F., Salazar, M.G., Learn, G.H., Morgan, C.J., Schumacher, J.E., Hraber, P., Giorgi, E.E., Bhattacharya, T., Korber, B.T., Perelson, A.S., Eron, J.J., Cohen, M.S., Hicks, C.B., Haynes, B.F., Markowitz, M., Keele, B.F., Hahn, B.H., Shaw, G.M.: High multiplicity infection by hiv-1 in men who have sex with men. PLoS Pathog **6**(5), 1000890 (2010)

[22] Rachinger, A., Groeneveld, P.H.P., van Assen, S., Lemey, P., Schuitemaker, H.: Time-measured phylogenies of gag, pol and env sequence data reveal the direction and time interval of hiv-1 transmission. AIDS **25**(8), 1035–9 (2011)

Table 1: Overview of the dataset composition.

risk group	# transmission chains	*gag*	*pol*	*env*
BC	7 (8)	3	-	7
HET	39 (50)	-	2	39
MSM	24 (26)	7	5	23
total	70 (84)	10	7	69

The number of analyzed transmission chains is given, with the total number of transmission events between brackets. The majority of the chains (60/70, 85.7%) represent 1 transmission event (range: 1 to 6). Sometimes, data of multiple genomic regions were available, augmenting the total number of datasets to n = 86. The number of chains is also given for each genomic region. Most data are available of the *env* region.

Table 2: Results of the fixed effects analyses.

	bottleneck size difference	δ	conditional effect size[3]	average bottleneck size (%)[4]
as model test[1]	no	3.4E-2	-0.48 (-3.15, 1.43)	96.5 (99.42)
all logistic[2]	no	1.6E-1	-3.15 (-15.92, 6.53)	96.2 (98.69)

[1] The demographic model of each dataset was specified according to the model selected by the PS and SS marginal likelihood estimators.

[2] In this analysis, all transmission chains were parametrized using the logistic growth model.

[3] This is the mean effect size and HPD interval (in ln units) of the predictor conditioned on inclusion of the effect in the model (i.e., $\beta_i|\delta_i = 1$).

[4] The average percentage of genetic diversity lost at transmission calculated using all datasets. The value between brackets was obtained after excluding the dataset with the smallest transmission bottleneck.

Table 3: Risk group evolutionary rate dataset characteristics.

subtype	risk group (# taxa)	date range	% male
A1	HET (44)	1986-2008	29.5
	IDU (23)	1997-2007	55.5
B	HET (54)	1996-2009	44.4
	MSM (73)	1983-2012	100
CRF01_AE	HET (77)	1990-2009	48.7[1]
	MSM (57)	2007-2011	100

[1] In these datasets there were a number of taxa with unknown gender origins. The reported male proportion does not take these into account.