AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 33, Number 8, 2017

DOI: 10.1089/aid.2016.0212

Short Communication: HIV-1 Transmission Networks Across South Korea

Mi Young Ahn,[1] Joel O. Wertheim,[2] Woo Joo Kim,[3] Shin-Woo Kim,[4] Jin Soo Lee,[5] Hea Won Ann,[1] Yongduk Jeon,[1] Jin Young Ahn,[1] Je Eun Song,[1] Dong Hyun Oh,[1] Yong Chan Kim,[1] Eun Jin Kim,[1] In Young Jung,[1] Moo Hyun Kim,[1] Wooyoung Jeong,[1] Su Jin Jeong,[1] Nam Su Ku,[1,6] June Myung Kim,[1,6] Davey M. Smith,[2] and Jun Yong Choi[1,6]

Abstract

Molecular epidemiology can help clarify the properties and dynamics of HIV-1 transmission networks in both global and regional scales. We studied 143 HIV-1-infected individuals recruited from four medical centers of three cities in South Korea between April 2013 and May 2014. HIV-1 *env* V3 sequence data were generated (337–793 bp) and analyzed using a pairwise distance-based clustering approach to infer putative transmission networks. Participants whose viruses were ≤2.0% divergent according to Tamura-Nei 93 genetic distance were defined as clustering. We collected demographic, risk, and clinical data and analyzed these data in relation to clustering. Among 143 participants, we identified nine putative transmission clusters of different sizes (range 2–4 participants). The reported risk factor of participants were concordant in only one network involving two participants, that is, both individuals reported homosexual sex as their risk factor. The participants in the other eight networks did not report concordant risk factors, although they were phylogenetically linked. About half of the participants refused to report their risk factor. Overall, molecular epidemiology provides more information to understand local transmission networks and the risks associated with these networks.

Keywords: HIV transmission network, HIV-1 *env* V3 sequence, South Korea

Introduction

The incidence of HIV infection has decreased in several countries, but the HIV epidemic is still spreading rapidly in many parts of world. In South Korea, the prevalence of HIV infection is still low (<0.03%), but the local epidemic has continued steadily since 1985 with around 800–1,300 new infections every year.[1] Interestingly, in a mathematical simulation, the HIV epidemic in South Korea was predicted to rapidly increase if new prevention measures are not undertaken.[2]

Understanding the underlying networks of HIV epidemics may be important to implement effective preventive measures.[3] In particular, techniques used in the molecular epidemiology can enhance our ability to characterize transmission networks of HIV infection in both global and regional scales,[4–6] and these approaches can be used to define clusters and spreaders.[5] Such identification may offer opportunities to target prevention interventions. To better understand the Korean HIV transmission network, we generated and used HIV-1 V3 sequence data from HIV-infected individuals receiving care at four medical centers. We then examined reported risk factors of these individuals to best describe the risks underlying the local transmission network.

Materials and Methods

Study population

We studied 250 HIV-infected participants who were prospectively recruited from four urban medical centers of three different cities in South Korea between April 2013 and May 2014. Samples (5 ml) of whole blood anticoagulated with EDTA were obtained from each participant. Plasma and peripheral blood mononuclear cells (PBMC) were separated by Ficoll-Paque density gradient centrifugation and aliquoted, frozen, and stored at −80°C until processing. Demographic and clinical data such as age, sex, self-reported HIV risk

[1]Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea.
[2]Department of Medicine, University of California, San Diego, La Jolla, California.
[3]Department of Internal Medicine, Korea University College of Medicine, Seoul, South Korea.
[4]Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, South Korea.
[5]Department of Internal Medicine, Inha University School of Medicine, Incheon, South Korea.
[6]AIDS Research Institute, Yonsei University College of Medicine, Seoul, South Korea.

exposures, and antiretroviral use were also collected. All studies were conducted in compliance with local institutional review board guidelines and with participants' written informed consent.

Sequencing of V3

HIV RNA was extracted from blood plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) and reverse-transcribed with the RETROscript Kit (Applied Biosystems, Foster City, CA) into complementary DNA (cDNA). Genomic DNA from PBMC was isolated using the QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's instructions. Nested polymerase chain reaction (PCR) of V3 was performed using 10 μl of diluted cDNA or cellular DNA template added to 40 μl of reaction mixture for the first round. The reaction mixture consisted of 5.0 μl of 10 × PCR buffer containing magnesium chloride and 1.0 μl of 10 nM dNTP Mix (GeneAmp; Applied Biosystems), 0.25 μl of Taq DNA polymerase (Roche Diagnostics, Indianapolis, IN), 31.75 μl of molecular grade water, and 1 μl of each of two 20 μM primers, V3-Fout (5′ GAGCCAATTCCCATAC ATTATTGT) and V3-Bout (5′GCCCATAGTGCTTCCTGC TGCTCCCAAGAACC). The 50 μl samples were heated to 94°C for 2 min and then subjected to 35 cycles of 30 s at 94°C followed by 30 s at 52°C followed by 60 s at 68°C. After this, the samples were heated to 68°C for 10 min and then held at 4°C until use. Second-round PCR used 5 μl of the first-round product as the template added to 45 μl of reaction mixture for a total volume of 50 μl. This reaction mixture consisted of the same reagents, but the volume of molecular grade water was increased to 36.75 μl. For this round, the primers used were V3-Fin (5′TGTGCCCCAGCTGGTTTTGCGAT) and V3-Bin (5′TATAATTCACTTCTCCAATTGTCC), and the thermal cycling parameters were 35 cycles of 30 s at 94°C followed by 30 s at 55°C followed by 60 s at 72°C. The PCR products from each sample were sequenced using Prism Dye terminator kits (ABI) on an ABI 3100 Genetic Analyzer.

Sequence analysis and transmission network inference

All sequences that were not duplicated, contaminated, or hypermutated by APOBEC were included, leaving 143 eligible sequences in the analysis to detect putative transmission clusters. Sequences were analyzed for genetic relatedness using a pairwise distance comparison using the Tamura-Nei 93 evolutionary model to correct for substitution biases and the unequal base composition found in HIV via HIV-TRACE (http://test.datamonkey.org/hivtrace).[6,7] Putative transmission clusters were inferred by establishing linkage when two sequences had a Tamura-Nei 93 genetic distance ≤2% separating them. Clusters were defined as connected components of the network comprising two or more nodes.

Epidemiological relatedness

We collected demographic and clinical data and reported exposure risks and compared those with genetic linkage data. Reported exposure risks included men who have sex with men (MSM), heterosexual contact, injecting drug use, receipt of blood/blood products, perinatal transmission, and other descriptions. Bisexuality was categorized as MSM. We conservatively considered that the risk exposures were concordant when all subjects within the pair reported specific risk exposure, and the reported risk were identical.

Results

A total of 67, 72, 65, and 54 participants were enrolled from hospital A, B, C, and D, respectively. From those, a total

Table 1. Characteristics of Participants Within Putative Transmission Clusters

Clusters	*Hospitals*	*Participants*	*Age (years)*	*Gender*	*Risk exposures*	*Year of diagnosis*
Pair 1	A	01-001	60	M	UK	2007
	B	02-057	37	M	UK	2012
Pair 2	A	01-031	42	M	UK	2009
	B	02-062	37	M	MSM	2011
Quartet	A	01-044	21	M	UK	2013
	A	01-045	68	M	MSM	2002
	A	01-059	24	M	UK	2013
	A	01-053	29	M	UK	2011
Pair 3	A	01-043	34	M	UK	2003
	B	02-037	43	M	MSM	2003
Pair 4	B	02-029	48	M	MSM	2013
	C	03-048	53	M	MSM	1992
Pair 5	B	02-049	66	M	MSM	2004
	B	02-051	63	F	Hetero	2004
Pair 6	B	02-047	40	M	MSM	2004
	C	03-036	56	M	Hetero	2005
Pair 7	C	03-012	23	M	UK	2012
	C	03-032	55	M	Hetero	2011
Pair 8	D	05-012	61	M	Hetero	2011
	D	05-015	44	M	UK	2005

M, male; UK, unknown; MSM, men who have sex with men; Hetero, heterosexual contact.

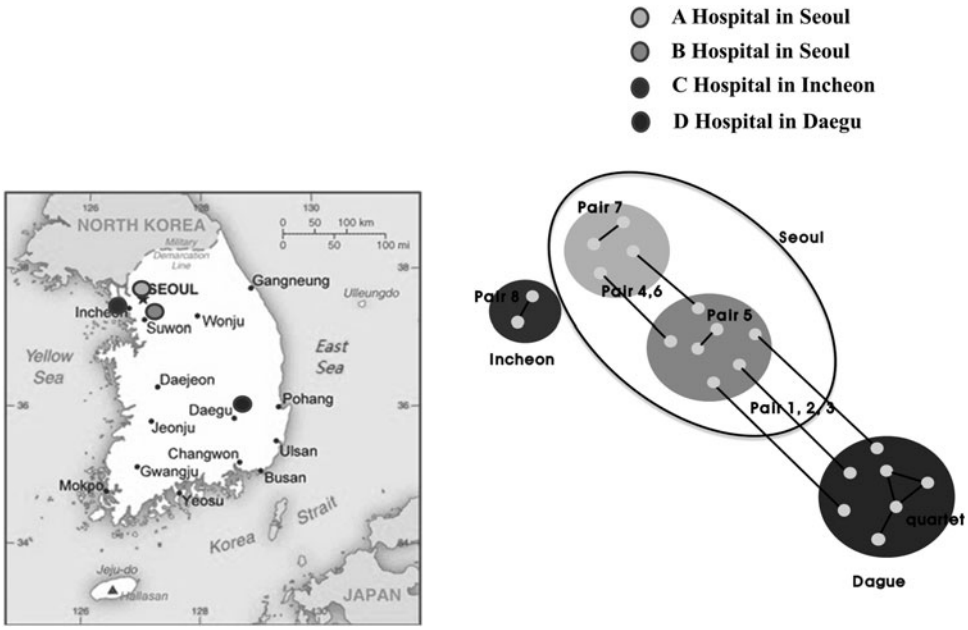

FIG. 1. Location of study sites and the putative transmission networks. *Shaded circles* represent each study sites and participants within identified putative transmission networks. The *black line* shows the potential transmission. Edge lengths and circle size are optimized for visual presentation and do not represent genetic distances between putative transmission partners. One *gray point* indicates one patient.

of 143 eligible V3 sequences were obtained. Of the 143 participants, 134 (93.7%) were male, and all were infected with HIV-1 subtype B excluding 4 participants (1 subtype C, 2 subtype CRF-1-AE, and 1 subtype G). Only 2 subjects were naive patients, and 141 were on antiretroviral treatments. One hundred eighteen subjects (82.5%) had suppressed viral loads (<20 copies/ml). We identified nine putative transmission clusters of different sizes (range 2–4 participants) consisting of 20 persons, that is, 14% of all participants were found in clusters. Eight clusters included two individuals per cluster (dyads), and one cluster included four individuals (quartet) (Table 1). Figure 1 shows putative transmission linkages and locations of each site. Among the nine clusters, four clusters included participants who were registered in different hospitals of different cities, and five clusters involved participants registered in different hospitals. The genetic distances of participants within each cluster are shown in Table 2. The phylogenetic tree shows the short genetic distance of each cluster (Fig. 2).

The reported risks were concordant in only one pair (pair 4). This pair reported their HIV risk as MSM. The genetic distance between their two sequences was close at 0.013 (Table 2). The participants in the other eight clusters consisted of 18 persons who did not report concordant risk factors, although their sequences were inferred to be closely related, but 9 persons in clusters (50.0%) did not report their risk exposures. Overall, among all enrolled individuals, 98/250 (39.2%) did not report their risk exposures.

Discussion

We applied a network approach to analyze Korean HIV-1 transmission clusters using generated V3 sequence data and defined clusters when sequences were inferred to be closely related (≤2.0% nucleotide genetic distance). This study is unique in that it used HIV-1 *env* V3 sequence data, rather than HIV-1 *pol* sequences, which is commonly used for drug-resistant testing.[8] In general, a 1%–2% genetic distance cutoff between *pol* sequences does performs well in balancing the identification of potential transmission partners, while not making excessive spurious connections.[7] This threshold is standard in the field and was selected based on previous work showing that within a monoinfected person, *pol* sequences typically do not diverge more than 1% from baseline over a decade.[9] The V3 region evolves much faster than *pol*, so we chose the upper limit as 2%. In particular, for the deeply sampled US HIV-1 subtype B epidemic, the genetic distance among two random HIV-1 *pol* sequences is between 4% and 7%,[6] while the mean genetic distance among two random V3 sequences is 13%. The mean genetic distance in the V3 data set was 0.117 substitutions/site, significantly lower than expected by chance. The maximum distance was 0.37 substitutions/site. Purifying selection in RNA viruses tells us that the selection of nucleotide substitution model does not affect the estimation of short genetic distances (e.g., 1%–2%).

Two participants of two pairs (pairs 1 and 8) had genetic distances of 0, meaning that the participants had basically the same V3 sequences; however, reported risk exposures were

Table 2. Genetic Distances Between Participants Within Clusters

	Subject 1	*Subject 2*	*Genetic distance*
Pair 1	01-001	02-057	0
Pair 2	01-031	02-062	0.019821
Quartet	01-044	01-045	0.002561
	01-045	01-059	0.015743
	01-044	01-059	0.016395
	01-053	01-059	0.016395
Pair 3	01-043	02-037	0.012702
Pair 4	02-029	03-048	0.013210
Pair 5	02-049	02-051	0.013416
Pair 6	02-047	03-036	0.012465
Pair 7	03-012	03-032	0.019368
Pair 8	05-012	05-015	0

Eight clusters included two individuals per cluster (dyads), and one cluster included four individuals (quartet).

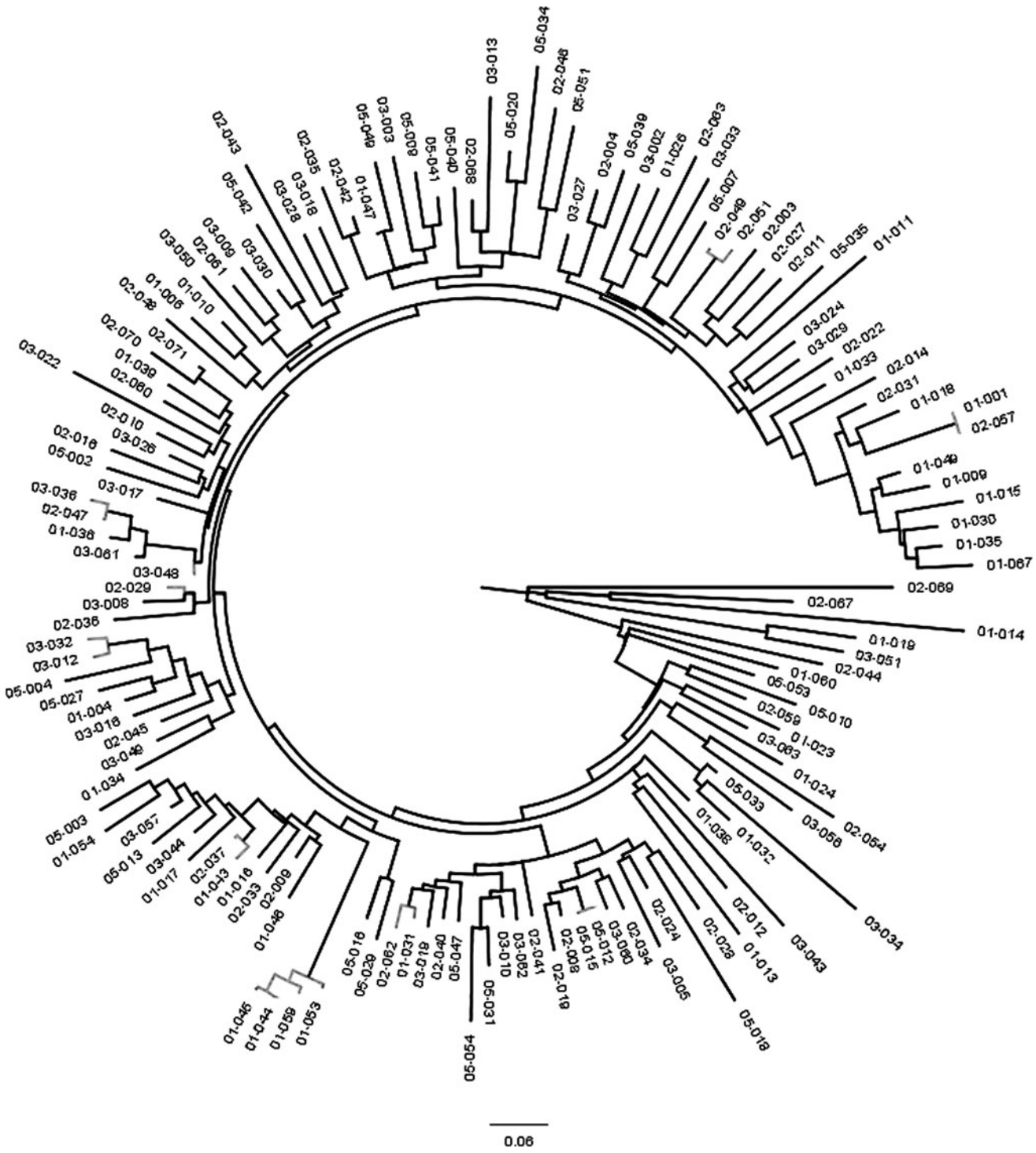

FIG. 2. A phylogenetic tree with V3 sequences of participants. *Gray line* indicates putative transmission networks that have genetic distance <2.

not concordant among these clustering individuals. With regard to the genetic distance of 0, we should consider the possibility of laboratory contamination, but genetic distance of ''0'' is not unexpected phenomenon. First, we resolved ambiguous nucleotides when calculating genetic distances, which means mixture of populations from which a transmitted variant arose would look identical to the original population using our method. Second, the short size of the sequences used here mean that identical sequences should not be unexpected if we got sequences from transmission pairs. In addition, about half of the participants refused to report their risk exposures. This presents a major obstacle to prevention efforts if reported risks are not available or not reliable. Although phylogenetic analysis may not be an

appropriate tool to identify direct transmissions,[10] a molecular epidemiologic approach to identify transmission networks might give us helpful information to understand local transmission networks and chances to implement effective preventive measures for targeted populations. We acknowledge that there are significant limitations to this approach including that the inferred transmission network is not always fully accurate and the presence of a genetic link between two individuals does not guarantee an HIV transmission event. In addition, selection bias may arise from studying participants who underwent sequencing and excluding those who did not; however, whereas differences are possible (e.g., participants who underwent sequencing may have had higher viral loads than those who did not), it is unclear how they would specifically affect the validity of inferred transmission networks. In addition, the small sample size mostly limits the generalization of this study. However, despite limitations, our study suggests molecular epidemiologic analysis for HIV-1 transmission networks could provide important information to understand local transmission networks and risks associated with these networks.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2005412), a Chronic Infectious Disease Cohort grant (4800-4859-304-260) from the Korea Centers for Disease Control and Prevention, BioNano Health-Guard Research Center funded by the Ministry of Science, ICT, and Future Planning of Korea as a Global Frontier Project (grant H-GUARD_2013M3A6B2078953), a grant from the Ministry of Health & Welfare, Republic of Korea (grant number: HI14C1324), a faculty research grant of Yonsei University College of Medicine (6-2015-0153) and NIH (grant P30 AI036214, AI100665, DA034978, AI036214 and K01AI110181).

Author Disclosure Statement

No competing financial interests exist.

References

1. CDC: Annual Report on the Notified HIV/AIDS in Korea. Statistical Annual Report of HIV/AIDS 2010.
2. Kim SB, Yoon M, Ku NS, *et al.*: Mathematical modeling of HIV prevention measures including pre-exposure prophylaxis on HIV incidence in South Korea. PLoS One 2014; 9:e90080.
3. Wang X, Wu Y, Mao L, *et al.*: Targeting HIV prevention based on molecular epidemiology among deeply sampled subnetworks of men who have sex with men. Clin Infect Dis 2015;61:1462–1468.
4. Little SJ, Kosakovsky Pond SL, Anderson CM, *et al.*: Using HIV networks to inform real time prevention interventions. PLoS One 2014;9:e98443.
5. Aldous JL, Pond SK, Poon A, *et al.*: Characterizing HIV transmission networks across the United States. Clin Infect Dis 2012;55:1135–1143.
6. Wertheim JO, Leigh Brown AJ, Hepler NL, *et al.*: The global transmission network of HIV-1. J Infect Dis 2014; 209:304–313.
7. Mehta SR, Wertheim JO, Brouwer KC, *et al.*: HIV transmission networks in the San Diego-Tijuana border region. EBioMedicine 2015;2:1456–1463.
8. Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJ: Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. Sci Rep 2016;6:39489.
9. Hightower GK, May SJ, Perez-Santiago J, *et al.*: HIV-1 clade B *pol* evolution following primary infection. PLoS One 2013;8:e68188.
10. Hecht FM, Wolf LE, Lo B: Lessons from an HIV transmission pair. J Infect Dis 2007;195:1239–1241.

Address correspondence to:
Jun Yong Choi
Department of Internal Medicine
Yonsei University College of Medicine
50 Yonsei-ro, Seodaemun-gu
Seoul 120-752
South Korea

E-mail: seran@yuhs.ac