Published in final edited form as:

Sci Transl Med. 2010 September ; 2(50): 501r1. doi:10.1126/scitranslmed.3001473.

Response to Comment on “The Origins of Sexually Transmitted HIV Among Men Who Have Sex with Men”

Davey Smith^{1,2,*†}, Wayne Delpont^{1,*}, David Butler¹, Susan Little¹, Douglas Richman^{1,2}, and Sergei Kosakovsky Pond¹

¹University of California, San Diego, La Jolla, CA 92093, USA

²Veterans Administration San Diego Healthcare System, San Diego, CA 92161, USA

Abstract

In this investigation, we evaluated explanations for the unusual degree of genetic diversity in HIV populations within the sampled seminal cell and plasma compartments observed in our previous study. These analyses included clonal sequencing of HIV DNA in peripheral blood mononuclear cells, ultradeep sequencing of HIV RNA in blood plasma, human leukocyte antigen (HLA) haplotyping of previously used samples, and a BLAST screen against both a local and public repository of HIV-1 sequences, and the investigations to determine whether these observations were secondary to contamination or artifact were unsuccessful. As there are very few HIV sequences from seminal cell tissues and transmission pairs described in the literature, future studies that evaluate more transmission pairs with sampling from multiple anatomic compartments and at multiple time points will most likely be required to resolve this controversy.

Experimental data made available through public repositories can greatly increase scientific progress by allowing others to reproduce published results and test other hypotheses. In the field of HIV research, the Los Alamos National Laboratory (LANL) HIV database (1) is the premier public resource for viral sequence and other data. The letter by Heath *et al.* (2) reanalyzes the sequence data from our study (3) and argues that “specimens that were mixed-up, mislabeled, or contaminated” could have led to “erroneous conclusions.” We appreciate the careful and thorough analyses presented in this report, and we thank the authors for sharing their results with us in advance. This has allowed us to perform a variety of follow-up investigations into possible contamination and sample mix-up, obtain new sequence and major histocompatibility complex (MHC) typing data, and correct several inaccurate labels of individual strains. In their letter, Heath *et al.* (2) argue that every seminal cell (SC) sequence in the original analysis was sampled from hosts other than any of the transmission sources in the study. Below, we evaluate this and other possible explanations of the observed phylogenetic separation of cell and plasma compartments in each of the source partners.

As identified by Heath *et al.* (2), some of the individual sequences in our original analysis were labeled incorrectly. We have carefully traced these sequences in our laboratory inventory system and determined that at least two problematic non-SC sequences (C810 and

[†]To whom correspondence should be addressed. davey@ucsd.edu.

*These authors contributed equally to this work.

Author contributions: D.S. and D.B. participated in designing the study, generating and analyzing the data, and writing the manuscript. W.D., S.K.P., S.L., and D.R. participated in designing the study, analyzing the data, and writing the manuscript.

Competing interests: D.R. has consulted for Chimerix, Gen-probe, Merck, Bristol-Myers Squibb, Gilead, Idenix, Monogram Biosciences, Tobira, Myriad, Biota, and Theraclone. D.S. has received grant support from Pfizer. The other authors have no competing interests.

C1632) were highly (mean, >20%) divergent from all other sequences from the same participant and anatomic compartment. Given the similarity of all other seminal and blood plasma sequences within participants, we concluded that these samples had been mislabeled. Accordingly, we have redesignated these two sequences as from an “unknown source/recipient.” Furthermore, because two sequences derived from SC HIV DNA (C958 and C949) were also extremely divergent from the SC sequence clade in the same participant (Table 1), we have removed these from the corrected analysis. In addition, seven samples in our original analyses were ascertained [through communication with B. Foley (4)] as attributed to an incorrect participant (Table 1). We corrected the labeling for four of these (C999, C992, C1008, and C1012) by tracing the samples in our laboratory inventory system. For the remaining three samples (SBP39, C1285, and C1270), we could not unambiguously resolve the labeling issue, so we removed them from our updated analysis and designated them as “unknown sources.” Although these mislabeled samples are a concern, they represent only a minority of the source SC sequences and do not affect the interpretation of the original results (Fig. 1).

The primary issue addressed by Heath *et al.* (2) is whether all the SC HIV-1 DNA sequences for all source patients were mislabeled or arose as a result of sample mix-up or contamination, and thereby led us to conclude erroneously that the HIV-1 RNA population in the seminal plasma was the origin of the transmitted virus (3). To evaluate the possibility of such a pervasive sample mix-up or contamination, we generated additional data and performed a series of follow-up analyses.

To see whether we could generate similar data from the SC fraction of semen from the source partners, we attempted to reamplify and sequence HIV C2-V3 from the remaining extracted SC DNA. We were unable to generate additional HIV polymerase chain reaction (PCR) product using the protocols from the original study (3), likely because of degradation of HIV DNA (stored at -20°C) during the 14 to 26 months after the original experiments.

We then performed ultradeep sequencing, as in (5), of HIV-1 RNA from blood plasma and clonal sequencing of HIV-1 DNA from peripheral blood mononuclear cells (PBMCs) for all source patients with two goals in mind. The first goal was to identify whether minority HIV-1 RNA variants in blood plasma matched HIV-1 DNA variants in SC, because this would confirm the presence of SC-like viruses in the blood and could be the result of mostly compartmentalized, divergent viral populations with limited movement of virus between cellular compartments. The second goal was to determine whether HIV-1 DNA viral populations in PBMCs were as divergent and compartmentalized as HIV-1 DNA in SCs, because this result would be most consistent with archived strains (perhaps from a different infection event) giving rise to divergent cell-associated clades in the phylogenetic analysis. For all source patients, both the PBMC and ultradeep sequences matched the viral variants previously sequenced from seminal and blood plasma, not the SC HIV-1 DNA. Because we could neither positively identify the origin of the SC viral sequences nor confirm strong differentiation in PBMCs, we proceeded to explore possible scenarios for mislabeling of samples or contamination.

If the observed pattern of strong compartmentalization between seminal plasma and SC virus within participants had been the result of sample mislabeling, it would have had to occur after the separation of seminal plasma from SCs. To evaluate whether SC samples had been mislabeled or mixed-up from their corresponding seminal plasma samples, we HLA (human leukocyte antigen)–haplotyped host DNA extracted from PBMC and SC samples. In all cases, the corresponding PBMC and SC haplotypes were identical. Because PBMC sequences were indistinguishable from blood and seminal plasma sequences within

participants, this result establishes that the SC and seminal plasma sequences were from the same person.

To further evaluate how multiple samples could be contaminated with a diverse viral population in only one of the three anatomic compartments tested, we evaluated for cross-contamination by ambient or historical HIV nucleic acid in our laboratory. We used BLAST to screen for putative contaminants from (i) a local repository of all HIV-1 *env* sequences ($n = 5200$) generated in our laboratory and (ii) the public sequence repositories (HIV LANL BLAST) (1). All SC sequences were HIV-1 subtype B of the C2-V3 region of *env*, and there were no significant matches either with sequences in the HIV LANL database or with other sequences generated in our laboratory during the 18-month period when the Butler *et al.* (3) study sequences were generated. To determine whether a virus from other biological samples in the lab could have contaminated our samples or amplifying reactions, we sequenced HIV-1 RNA extracted from the blood plasma of study subjects in our local cohort who had not previously been sequenced. To our surprise, we discovered that sequences from source partner A's SCs were 99% identical to sequences from another study subject. There is no record that this study subject's virus had been PCR-amplified or sequenced before or during the time the Butler *et al.* (3) study was being conducted; however, we cannot definitively rule out this contamination possibility.

In summary, we agree that the apparent lack of relatedness between samples from plasma and seminal cellular compartment in all source patients, as noted by Heath *et al.* (2), is highly unusual and the default explanation is that SC "specimens...were mixed-up, mislabeled, or contaminated." However, despite our best efforts and performance of every due diligence test that we could think of, we found no direct evidence for repeated or batched contamination, mix-up, or mislabeling. Nevertheless, however unlikely, contamination remains a possible explanation for our results. On the other hand, we are not prepared to discount alternative biological explanations. For source A, we discovered that another subject from the same cohort had 99% identical C2-V3 *env* sequences. This could result from superinfection of source A by this subject (or by another individual in the same HIV transmission network) (6) or by contamination (7). A potential source of contamination could have been PCR-amplified virus from the other subject, but no record of such amplification exists in our laboratory's comprehensive inventory system. Even if we conservatively discard both transmissions from source A, there remain three transmission pairs where no direct or indirect evidence for contamination has been identified. As proposed by Heath *et al.* (2), the original results could be consistent with multiple infections in the source with viral compartmentalization in SC tissues. Although not discussed in the original study (3), all source partners had reported a large number of partners in the 6 months preceding their enrollment into the study (mean, 8; range, 2 to 30), and generally, the study cohort has a high rate of superinfection (8), which is common in other cohortsof men who have sex with men (9). Because there are very few HIV sequences from SC tissues and transmission pairs described in the literature, it remains unclear whether the Butler *et al.* (3) result represents a rare and previously unrecognized biological phenomenon, perhaps attributable to the remarkably high rate of multiple infections in the study cohort. Furthermore, the lack of phylogenetic clustering among multiple samples of source SCs, for instance, in source D [Fig. 1 in (2)], does not prove that these samples have been derived from different patients. Indeed, there is little power to resolve most branching orders in the tree built on short C2-V3 sequence reads. The bootstrap support of source D SSC sequences with unrelated subtype B sequences is very low (Fig. 2). Currently, we do not have sufficient knowledge of temporal dynamics of HIV diversity within seminal compartments to attribute our results to generic unknown contamination. Ultimately, this debate will be resolved by the sequencing and characterization of additional transmission pairs. We have continued to sample transmission pairs for multiple compartments and at multiple time

points with the purpose of confirming or refuting our original results and will present these data as soon as they become available.

Acknowledgments

We thank B. Foley and R. Swanstrom for their insights and advice and S. Mehta, S. Gianella Weibel, and M. Pacold for additional technical assistance.

Funding: This work was supported by grants from the NIH (AI69432, AI043638, MH62512, MH083552, AI077304, AI36214, AI047745, AI7462, AI080353, and AI57167); a Joint Division of Mathematical Sciences/ National Institute of General Medical Sciences Mathematical Biology Initiative Grant NSF-0714991; the University of California University-Wide AIDS Research Program (grant number IS02-SD-701); University of California, San Diego Center for AIDS Research/National Institute of Allergy and Infectious Diseases Developmental Award AI36214 (to S.K.P. and W.D.); and the Veterans Affairs Healthcare System (San Diego, California).

REFERENCES AND NOTES

1. Los Alamos National Laboratory. [accessed 2010]. <http://www.hiv.lanl.gov>
2. Heath L, Frenkel LM, Foley BT, Mullins JI. The origins of sexually transmitted HIV among men who have sex with men. *Sci. Transl. Med.* 2010; 2:50le1. Comment on.
3. Foley BT. personal communication. March.2010
4. Butler DM, Delpont W, Kosakovsky Pond SL, Lakdawala MK, Cheng PM, Little SJ, Richman DD, Smith DM. The origins of sexually transmitted HIV among men who have sex with men. *Sci. Transl. Med.* 2010; 2:18re1.
5. Smith, DM. paper presented at the 17th International Conference on HIV Dynamics and Evolution; Monterey, CA. 25 to 28 April 2010;
6. Smith DM, May SJ, Tweeten S, Drumright L, Pacold ME, Kosakovsky Pond SL, Pesano RL, Lie YS, Richman DD, Frost SD, Woelk CH, Little SJ. A public health model for the molecular surveillance of HIV transmission in San Diego, California. *AIDS.* 2009; 23:225–232. [PubMed: 19098493]
7. Frenkel LM, Mullins JI, Learn GH, Arcuino L, Herring BL, Kalish ML, Steketee RW, Thea DM, Nichols JE, Liu SL, Harmache A, He X, Muthui D, Madan A, Hood L, Haase AT, Zupancic M, Staskus K, Wolinsky S, Krogstad P, Zhao J, Chen I, Koup R, Ho D, Korber B, Apple RJ, Coombs RW, Pahwa S, Roberts NJ Jr. Genetic evaluation of suspected cases of transient HIV-1 infection of infants. *Science.* 1998; 280:1073–1077. [PubMed: 9582120]
8. Smith DM, Wong JK, Hightower GK, Ignacio CC, Koelsch KK, Daar ES, Richman DD, Little SJ. Incidence of HIV superinfection following primary infection. *JAMA.* 2004; 292:1177–1178. [PubMed: 15353529]
9. Sidat MM, Mijch AM, Lewin SR, Hoy JF, Hocking J, Fairley CK. Incidence of putative HIV superinfection and sexual practices among HIV-infected men who have sex with men. *Sex. Health.* 2008; 5:61–67. [PubMed: 18361856]
10. Smith D, Delpont W, Butler D, Little S, Richman D, Pond S. Kosakovsky. *Sci. Transl. Med.* 2010; 2:50lr1. Response to comment on “The origins of sexually transmitted HIV among men who have sex with men.”.

Fig. 1. (A to F) Phylogenetic analysis of source and recipient viral sequences. Maximum likelihood phylogenies were generated for each transmission pair as in Butler *et al.* (3), but with problematic sequences removed. Taxa are colored by anatomic compartment and labeled with GenBank accession numbers. RBP, recipient HIV RNA in blood plasma; SBP, source HIV RNA in blood plasma; SSP, source HIV RNA in seminal plasma; SSC, source SC-associated virus.

Fig. 2. Phylogenetic analysis of all source and recipient viral sequences. A neighbor-joining phylogeny was generated and percent node support was determined with bootstrap resampling. Magenta branches are 50 randomly selected HIV-1 subtype B sequences from the LANL HIV database (1).

Table 1

Mislabeled and potentially mislabeled individual sequences. SBP, source blood plasma HIV-1 RNA; RBP, recipient blood plasma HIV-1 RNA; SSC, source SC HIV-1 DNA

	Butler <i>et al.</i> , 2010 label	Corrected label	GenBank accession number
Identified by Heath <i>et al.</i>			
C810	Recipient D RBP	Unknown	GU597092
C1632	Source F SBP	Unknown	GU597193
C958	Source A SSC	Unknown	GU597266
C949	Source A SSC	Unknown	GU597267
Identified before Heath <i>et al.</i>			
C999	Source E SSC	Source A SSC	GU597238
C992	Source E SSC	Source A SSC	GU597240
C1008	Source E SSC	Source A SSC	GU597241
C1012	Source E SSC	Source A SSC	GU597239
SBP39	Source E SBP	Unknown	GU597180
C1285	Source F SSC	Unknown	GU597256
C1270	Source F SSC	Unknown	GU597257