

## Perspectives

# Inferring HIV Transmission Dynamics from Phylogenetic Sequence Relationships

Christopher D. Pilcher\*, Joseph K. Wong, Satish K. Pillai

Despite the range of resources directed at understanding the HIV pandemic over the past 25 years, surprisingly little is known about how HIV infection spreads through populations. Unlike some other infectious diseases, acute infection with HIV is difficult to identify. HIV disease most often manifests years after the transmission event. Together with the special challenges involved in determining exposures related to sexual behavior or drug use, all of these factors have made it difficult to apply the tools of traditional epidemiologic investigation. Recent antibody testing strategies to identify incident HIV for surveillance programs have met with limited success [1]. Key questions that remain unanswered by empirical data include the role of acute infections in sustaining the current pandemic, and the effects of antiretroviral treatment programs on transmission of drug-resistant and drug-susceptible strains of HIV. Without really understanding how HIV spreads, it is difficult to optimize prevention or control strategies.

As effective anti-HIV therapies emerged over the past decade, clinical care and surveillance programs have increasingly emphasized the importance of testing for resistance to antiretroviral drugs. This most commonly involves sequencing of viral genes for resistance mutations. The rapid expansion of this HIV genotyping has predictably resulted in creation of vast databases that now contain viral sequence information. The new study by Andrew Leigh Brown and colleagues in this issue of *PLoS Medicine* [2] shows that modern analytic tools may yield important new insights into HIV transmission dynamics from the information routinely collected in such sequence databases.

The Perspectives section is for experts to discuss the clinical practice or public health implications of a published article that is freely available online.

## Linked Research Article

This Perspective discusses the following new study published in *PLoS Medicine*:

Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ (2008) Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 5(3): e50. doi:10.1371/journal.pmed.0050050

Using viral genotype data from HIV drug resistance testing at a London clinic, Andrew Leigh Brown and colleagues derive the structure of the transmission network through phylogenetic analysis.

## HIV “Phylodynamics” for the Study of Local Epidemiology

Leigh Brown and colleagues were interested in better understanding the epidemiology of HIV among men who have sex with men in London. To this end, they obtained access to a relatively large convenience sample of HIV *pol* sequences (see Glossary) obtained through the routine testing of 2,126 unique HIV-infected patients served by a large university medical center in London. They used a “phylodynamic” approach, an interdisciplinary blend of immunodynamics, epidemiology, and evolutionary biology, to infer the short-term dynamics of HIV transmission in the base population from relationships among sequences in their study sample.

The authors initially applied a viral genetic relatedness cutoff to filter the data down to a computationally manageable subset of 402 HIV-infected individuals that exhibited at least one other close sequence relative in the study population. Nine large putative transmission clusters were identified within this subset of protease and reverse transcriptase sequence data on the basis of genetic (Hamming) distance. The presence of these transmission clusters was subsequently independently verified using Bayesian Markov chain Monte

Carlo phylogenetic methodology. The authors then used a “relaxed clock” approach to generate time-scaled phylogenies of these data, to infer the timing and distribution of transmission events within the 88 sequences contained in the six clusters that were large enough for analysis.

While components of the methodology were previously established and applied in other contexts, the results of this first successful application of phylodynamics to HIV sequence data-mining are themselves noteworthy for several particular reasons. First, the internode distances within the study’s time-scaled phylogenies were surprisingly short—in more than a quarter of cases, transmission events appear to have occurred fewer than six months after infection. Second, a substantial majority of the transmissions inferred to have taken place in the clusters were concentrated in a well-defined five-year period, bounded by periods of less frequent transmission. Together, the phylodynamic data

**Funding:** CDP is supported in part by National Institutes of Health/National Institute of Allergy and Infectious Diseases grant R01 MH068686. JKW and SKP are supported by National Institutes of Health grant R01 NS051132 and the Department of Veterans Affairs. The funders played no role in the preparation of the article.

**Competing Interests:** The authors have declared that no competing interests exist.

**Citation:** Pilcher CD, Wong JK, Pillai SK (2008) Inferring HIV transmission dynamics from phylogenetic sequence relationships. *PLoS Med* 5(3): e69. doi:10.1371/journal.pmed.0050069

**Copyright:** © 2008 Pilcher et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Christopher D. Pilcher, Joseph K. Wong, and Satish K. Pillai are at the University of California San Francisco, San Francisco, California, United States of America. Joseph K. Wong and Satish K. Pillai are also at the San Francisco VA Medical Center, San Francisco, California, United States of America.

\* To whom correspondence should be addressed. E-mail: cpilcher@php.ucsf.edu

suggest that the (sexual) transmission of HIV in London over the previous decade may have occurred not as a slow and steady process, but rather via discrete outbreaks fueled in part by efficient transmission during acute HIV infection.

### Differences from Previous Studies

Phylogenetic sequence analysis has been used extensively in HIV epidemiology. These data are commonly used to support the identity of supposed “transmission pairs” for purposes of contact investigation [3], translational biological studies [4], and epidemiologic studies in which HIV transmission is an outcome [5]. Looking at larger sequence databases, a number of investigators have taken the clustering outcome as evidence of individual membership in a contact network or as an (indirect) marker of infectivity. Their studies have correlated clustering with acute disease stage [6–8], viral factors [9], risk behaviors [7,10], and even geography [10]. The present study is distinguished from these reports by its focus on the internal architecture of the sequence clusters. Leigh Brown and colleagues’ ability to study internal cluster structure clearly depends on access to large numbers of clustered sequences (which might relate in turn to either the structure of underlying contact networks or to the density of population sampling).

### Implications for Public Health and Clinical Practice

If application of Leigh Brown and colleagues’ phylodynamic methods to HIV can be further validated and their results confirmed by additional investigators, the finding that HIV is frequently transmitted through discrete outbreaks would suggest the need for a stronger emphasis on outbreak detection and network intervention/outbreak control strategies [11]. These strategies are currently used for other diseases, such as syphilis and tuberculosis. In this context, it is worth noting that sequence data-mining techniques can be as easily misused as used properly [12]. Guidelines are needed to clarify individual privacy rights and provide a legal framework for dealing with such sequence data that balances patient autonomy with scientific and public health objectives. Until then, exceptional

caution should be used in dealing with phylogenetic/dynamic associations at the individual level.

Most immediately, the ability to illustrate epidemic dynamics through the analysis of phylogenetic sequence information should encourage surveillance and prevention researchers to explore sequence databases with renewed vigor. With the revision of guidelines encouraging more frequent resistance testing of newly diagnosed patients [13], and the new creation of sequence databases worldwide, hopefully the number of populations with the high-density sampling necessary for phylodynamic analysis may be increasing. What is occurring globally, in diverse settings, with the introduction of antiretroviral treatment programs? To what degree is transmission efficiency affected

### Glossary

**Hamming distance:** The number of nucleotide differences between two genetic sequences.

**HIV *pol* sequence:** The HIV *pol* gene encodes all three of the viral enzymes (protease, reverse transcriptase, and integrase), and is the principal target of antiretroviral therapy. Data used by Leigh Brown and colleagues included the protease and partial reverse transcriptase sequences.

**Internode distance:** Each node in a phylogenetic tree represents the most common recent ancestor of its descendants. Within HIV phylogenies that include a single sequence representative per infected individual, the distance between each most common recent ancestor and the previous node estimates the upper bound of time between transmission events.

**Markov chain Monte Carlo methods:** A class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution.

**Relaxed clock approach:** Using extent of sequence change to infer the time interval between related viral variants (molecular clock hypothesis), taking into account rate variation across lineages to obtain better estimates of divergence times.

by drug resistance, and how will this affect future treatment options? Do the dynamics of devastating epidemics in sub-Saharan Africa or Eastern Europe differ in some fundamental way from those in the most developed countries? The provocative data from Leigh Brown and colleagues suggest an outbreak model for London’s community of men who have sex with men; similar and complementary investigations in diverse settings should clarify the actual need for new global HIV control strategies. ■

### References

1. McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, et al. (2006) Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Res Hum Retroviruses* 22: 945-952.
2. Lewis F, Hughes CJ, Rambaut A, Pozniak A, Leigh Brown AJ (2008) Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 5: e50. doi:10.1371/journal.pmed.0050050
3. Blick G, Kagan RM, Coakley E, Petropoulos C, Maroldo L, et al. (2007) The probable source of both the primary multidrug-resistant (MDR) HIV-1 strain found in a patient with rapid progression to AIDS and a second recombinant MDR strain found in a chronically HIV-1-infected patient. *J Infect Dis* 195: 1250-1259.
4. Zhu T, Wang N, Carr A, Nam DS, Moor-Jankowski R, et al. (1996). Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. *J Virol* 70: 3098-3107.
5. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, et al. (2005) Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 191: 1403-1409.
6. Yerly S, Vora S, Rizzardi P, Chave JP, Vernazza PL, et al. (2001) Swiss HIV Cohort Study. Acute HIV infection: Impact on the spread of HIV and transmission of drug resistance. *AIDS* 15: 2287-2292.
7. Pao D, Fisher M, Hué S, Dean G, Murphy G, et al. (2005) Transmission of HIV-1 during primary infection: Relationship to sexual risk and sexually transmitted infections. *AIDS* 19: 85-90.
8. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, et al. (2007) Quebec Primary HIV Infection Study Group. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* 195: 951-959.
9. Lindström A, Ohliss A, Huigen M, Nijhuis M, Berglund T, et al. (2006) HIV-1 transmission cluster with M41L ‘singleton’ mutation and decreased transmission of resistance in newly diagnosed Swedish homosexual men. *Antivir Ther* 11: 1031-1039.
10. Frost S, McCoy S, Hicks C, Williams D, Eron J, et al. (2007) Tracking molecular epidemiology in North Carolina, USA: The screening and tracing active transmission model [abstract 240]. 14th Conference on Retroviruses and Opportunistic Infections; 25-28 February 2007; Los Angeles, California, United States of America. Available: <http://www.retroconference.org/2007/Abstracts/30268.htm>. Accessed 8 February 2008.
11. Pilcher CD, Eaton L, Kalichman S, Bisol C, de Souza Rda S (2006). Approaching “HIV

elimination": Interventions for acute HIV infection. *Curr HIV/AIDS Rep* 3: 160-168.

12. Hecht FM, Wolf LE, Lo B (2007) Lessons from an HIV transmission pair. *J Infect Dis* 195: 1239-1241.

13. US Department of Health and Human Services (2007) Guidelines for the use of antiretroviral agents in hiv-infected adults and adolescents. DHHS Panel on Antiretroviral Guidelines for Adults and

Adolescents—A Working Group of the Office of AIDS Research Advisory Council (OARAC). Available: <http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL000721.pdf>. Accessed 8 February 2008.

