Sequencing Applications in HIV Forensics

Steve Scherer
Human Genome Sequencing Center
Department of Molecular and Human Genetics
Baylor College of Medicine

ABRF 2013

Application of Next Generation Sequencing Technologies for Whole Transcriptome and Genome Analysis Workshop
Phylogenetic Linkage Cases

Establishing Linkage:

• The Florida Dentist – 6 patient samples were more closely related to each other than unrelated controls
• The French Surgeon
• The French Nurse
• Criminal cases of HIV transmission in Sweden, Australia, Belgium, Denmark, Germany and Scotland

Established Non-Linkage:

• Baltimore Surgeon
• UK obstetrician/gynecologist
Molecular Forensics

• National: 2001 Anthrax in the Mail Case – Amerithrax
  - Lengthy FBI/DOJ investigation – Bruce Ivins
  - Highly controversial
  - National Academy of Sciences Review: “impossible to reach any definitive conclusion about the origins of the anthrax in the letters, based solely on the available scientific evidence”

• International: The Poison Umbrella Case
  - Georgi Markov: Bulgarian activist, playwright & satirist was working for the BBC World Service in London
Ricin and the umbrella murder

Thursday, October 23, 2003 Posted: 2:27 AM EDT (06:27 GMT)

LONDON, England -- it was one of the most notorious acts of assassination carried out during the Cold War.

Bulgarian dissident Georgi Markov was killed by a poison dart filled with ricin and fired from an umbrella in London in 1978.

Markov, a communist defector working for the BBC World Service, left his office at Bush House in the UK capital on September 11 and walked across Waterloo Bridge to take the train home to Clapham in south-west London.
DNA Testing in the Judicial System

• In use since the mid ‘80’s

• Usually used to link perpetrators to violent crime scenes

• Generally stable allowing the use of polymorphic marker DNA fingerprinting - error correcting machinery
Dynamic evolution of HIV

Individuals infected with HIV-1 contain a dynamically evolving population of related genomes

• Viral expansion
  – Mutation rate
  – Recombination rate
  – Production of $10^8$ to $10^{10}$ virions per day

• Lineage extinction
  – Non-replicating virions
  – Immune system
  – Drug therapy

Phylogenetics in Forensics

Viral dynamics limit use the common practice of matching DNA profiles.

Phylogenetic methods are ideally suited for determining the HIV pattern of descent in suspected transmission cases.

In support of the *a priori* hypothesis, HIV forensics can identify case samples to be “more closely related” than to unrelated samples.

In opposition of the *a priori* hypothesis, HIV forensics can show case samples to be unrelated.
HIV forensics

State of Louisiana v. Richard J. Schmidt
Patient → Trahan

State of Washington v. Anthony E. Whitfield
Whitfield → 5 partners

State of Texas v. Philippe Padieu
Padieu → 6 partners

Metzker et al. (2002) PNAS 99: 14292-14297

Scaduto et al. (2010) PNAS 107: 21242-21247
HIV genes: pol and env

Methods involved were:
- Fractionation of PBMCs
- Isolation of genomic DNA
- PCR and cloning
- Sanger sequencing
- Multiple sequence alignments
State of Louisiana v. Richard J. Schmidt
Patient → Victim
Prosecution argued successfully:

Methods used for:
- DNA isolation
- PCR
- DNA sequencing
- Phylogenetic analysis of HIV-1 positive samples

Met the standards of judicial evidence admissibility

Due to:
- Are subject to empirical testing
- Can be assessed for error
- Subject to peer review and publication
- Generally accepted in the scientific community
Case Facts

• August 4, 1994 – a Lafayette, LA gastroenterologist created a cocktail of two blood samples; one from a man infected with HIV-1 and a second from a patient infected with Hep-C, then infected his former girlfriend by intramuscular injection
• Victim had 7 sexual relationships between 1984 and 1995; all tested negative for HIV
• Victim was a nurse in the Lafayette area; no documented needle sticks; one HIV + saliva splash; tested negative
• Victim donated blood and tested negative 10/92, 5/93 & 4/94
• Victim tested positive in January 1995 – accused the physician
• Police raid discovered potential source blood draw record
Testing Protocol

1. Data was derived from two separate labs and two independent blood draws; the second lab was blinded.
2. Samples were tested one at a time – lab scrub between samples.
3. PCR amplified env and RT genes
   - exhibit different biological functions
   - subject to different selective pressures
   - known to undergo different rates of evolution.
4. 50 molecular clones from sample were sequenced to further delineate env genetic diversity.
5. Geographic HIV-1+ controls were recruited from the Lafayette metro area based on previous studies demonstrating geographic subtype stratification of HIV-1 sequences.
Table 1. Summary of LA control group sample sources

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Percentage</th>
<th>Count (Cases/Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homosexual</td>
<td>57%</td>
<td>(16/28)</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>18%</td>
<td>(5/28)</td>
</tr>
<tr>
<td>Blood transfusion</td>
<td>11%</td>
<td>(3/28)</td>
</tr>
<tr>
<td>Bi-sexual</td>
<td>7%</td>
<td>(2/28)</td>
</tr>
<tr>
<td>IV drug user</td>
<td>4%</td>
<td>(1/28)</td>
</tr>
<tr>
<td>Sharps</td>
<td>4%</td>
<td>(1/28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date infected</th>
<th>Percentage</th>
<th>Count (Cases/Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983–1989</td>
<td>57%</td>
<td>(16/28)</td>
</tr>
<tr>
<td>1990–1992</td>
<td>36%</td>
<td>(10/28)</td>
</tr>
<tr>
<td>Unknown</td>
<td>7%</td>
<td>(2/28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CD4⁺ cell counts</th>
<th>Percentage</th>
<th>Count (Cases/Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>36%</td>
<td>(10/28)</td>
</tr>
<tr>
<td>200–500</td>
<td>25%</td>
<td>(7/28)</td>
</tr>
<tr>
<td>&lt;200</td>
<td>25%</td>
<td>(7/28)</td>
</tr>
<tr>
<td>ND</td>
<td>14%</td>
<td>(4/28)</td>
</tr>
</tbody>
</table>

Risk factors and dates infected were obtained by anonymous questionnaire. ND, not determined.
Phylogenetic Analyses

For the case: parsimony and minimum evolution using maximum-likelihood distances

Subsequently: Markov-chain Monte Carlo Bayesian Analysis based on a General-Time-Reversible model of sequence evolution, with $\gamma$-distributed rate heterogeneity among sites and a calculated proportion of invariable sites
Table 2. Means and 95% confidence intervals for parameters of the GTR + $\Gamma$ + I model for gp120 sequences

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>C–T substitution rate</td>
<td>5.03</td>
<td>3.60–7.03</td>
</tr>
<tr>
<td>C–G substitution rate</td>
<td>0.97</td>
<td>0.57–1.54</td>
</tr>
<tr>
<td>A–T substitution rate</td>
<td>0.75</td>
<td>0.52–1.07</td>
</tr>
<tr>
<td>A–G substitution rate</td>
<td>3.87</td>
<td>2.91–5.10</td>
</tr>
<tr>
<td>A–C substitution rate</td>
<td>2.34</td>
<td>1.60–3.34</td>
</tr>
<tr>
<td>Frequency of A</td>
<td>0.40</td>
<td>0.37–0.43</td>
</tr>
<tr>
<td>Frequency of C</td>
<td>0.15</td>
<td>0.13–0.17</td>
</tr>
<tr>
<td>Frequency of G</td>
<td>0.23</td>
<td>0.21–0.25</td>
</tr>
<tr>
<td>Frequency of T</td>
<td>0.22</td>
<td>0.20–0.25</td>
</tr>
<tr>
<td>$\alpha$ (shape of $\Gamma$ distribution)</td>
<td>0.53</td>
<td>0.43–0.68</td>
</tr>
<tr>
<td>Proportion of invariable sites</td>
<td>0.08</td>
<td>0.01–0.18</td>
</tr>
</tbody>
</table>

Data based on MCMC sampling (25). The rate of all substitution classes is shown relative to that of the G–T substitution class.
Table 3. Means and 95% confidence intervals for parameters of the GTR + Γ + I model for the RT sequences

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>C–T substitution rate</td>
<td>110.36</td>
<td>23.04–195.53</td>
</tr>
<tr>
<td>C–G substitution rate</td>
<td>17.59</td>
<td>2.82–42.02</td>
</tr>
<tr>
<td>A–T substitution rate</td>
<td>7.62</td>
<td>1.34–17.32</td>
</tr>
<tr>
<td>A–G substitution rate</td>
<td>83.01</td>
<td>16.29–171.17</td>
</tr>
<tr>
<td>A–C substitution rate</td>
<td>16.60</td>
<td>3.41–35.62</td>
</tr>
<tr>
<td>Frequency of A</td>
<td>0.40</td>
<td>0.36–0.43</td>
</tr>
<tr>
<td>Frequency of C</td>
<td>0.17</td>
<td>0.14–0.19</td>
</tr>
<tr>
<td>Frequency of G</td>
<td>0.20</td>
<td>0.17–0.23</td>
</tr>
<tr>
<td>Frequency of T</td>
<td>0.23</td>
<td>0.20–0.26</td>
</tr>
<tr>
<td>α (shape of Γ distribution)</td>
<td>0.94</td>
<td>0.38–1.94</td>
</tr>
<tr>
<td>Proportion of invariable sites</td>
<td>0.50</td>
<td>0.29–0.63</td>
</tr>
</tbody>
</table>

Data are based on MCMC sampling (25). The rate of all substitution classes is shown relative to the rate of the G–T substitution class.
Phylogenetic analysis of the gp120 region using a minimum evolution criterion and maximum likelihood distances assuming an HKY+Γ model of evolution.

Metzker M L et al. PNAS 2002;99:14292-14297
Phylogenetic analysis of the RT region; details of the analysis are the same as for Fig. 1.

Metzker M L et al. PNAS 2002;99:14292-14297

©2002 by National Academy of Sciences
Reverse transcriptase (RT) alignment of predicted amino acid residues
Results

For RT: Patient sequences were found to be paraphyletic with respect to victim sequences (victim sequences nested within patient sequences) by parsimony, minimum evolution, and Bayesian analyses.

For env gp120: Not as strong; supports a weak mono-phyletic grouping. Probably due faster evolution and strong immuno-selection.

Defendant was found guilty of attempted second degree murder; Appeal rejected by the Louisiana Supreme Court; U.S. Supreme Court declined to the case.
Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences

Diane I. Scaduto\textsuperscript{a,b}, Jeremy M. Brown\textsuperscript{c,1}, Wade C. Haaland\textsuperscript{a,b}, Derrick J. Zwicki\textsuperscript{c,2}, David M. Hillis\textsuperscript{c,3}, and Michael L. Metzker\textsuperscript{a,b,d}

Author Affiliations

Author Notes

Contributed by David M. Hillis, October 20, 2010 (sent for review September 22, 2010)
State of Washington v. Anthony E. Whitfield
Whitfield → 5 partners

State of Texas v. Philippe Padieu
Padieu → 6 partners
Direction of transmission  
(source → recipient)

Providing evidence for the direction of transmission would further strengthen the *a priori* hypothesis.

Genetic bottleneck during transmission

• Paraphyly: Evidence for direction of transmission

Study design:

• identities of case subjects were blinded to investigators

• case sample handling were separated both temporally and spatially to eliminate the possibility of cross contamination

• case allegations were multiple transmissions from a single source
Washington Case

• Defendant allegedly learned of his HIV positive status in April 1992; 17 partners were exposed between 1999 and 2004; 5 tested positive between 2002 and 2004
• These 6 samples formed the basis of the *a priori* hypothesis of transmission from one source to multiple recipients

Texas Case

• Defendant allegedly learned of his HIV positive status in September 2005; 6 partners tested positive between April 2006 and March 2007
• These 7 samples formed the basis of the *a priori* hypothesis of transmission from one source to multiple recipients
Washington case: ML tree for the pol gene dataset using BLAST-selected GenBank controls.

Scaduto D I et al. PNAS 2010;107:21242-21247
Washington case: ML tree for the env gene dataset using BLAST-selected GenBank controls.

Scaduto D I et al. PNAS 2010;107:21242-21247
CC01 exhibited a paraphyletic relationship to all CC case sequences
- Bayesian posterior probabilities (1.00)
- ML bootstrapping proportions (0.98)

Red circle represents the most recent common ancestor of sequences from CC01

Scaduto et al. (2010) *PNAS* 107: 21242-21247
Texas case: *env* tree

CC01 exhibited a paraphyletic relationship to all CC case sequences but CC05

- Bayesian posterior probabilities (1.00)
- ML bootstrapping proportions (1.00)

Red circle represents the most recent common ancestor of sequences from CC01

Scaduto et al. (2010) *PNAS* **107**: 21242-21247
Breaking the code

For the WA case, we inferred that sample WA04 was the source (i.e., index case)

   At trial, the identity of sample WA04 was revealed to be that of Anthony E. Whitfield

For the TX case, we inferred that sample CC01 was the index case

   At trial, the identity of sample CC01 was revealed to be that of Philippe Padieu

Scaduto et al. (2010) PNAS 107: 21242-21247
Texas case: env tree

CC07 exhibited a paraphyletic relationship to several CC01 and all CC03 sequences

- Bayesian posterior probabilities (1.00)
- ML bootstrapping proportions (0.79)

Scaduto et al. (2010) PNAS 107: 21242-21247
Testing for Recombination in env

Used a maximum likelihood method that detects recombination breakpoints using a hidden Markov model.*

If two trees are allowed across the sequence, it estimated a recombination structure like this:

Divergent Signals Between Halves

1\textsuperscript{st} half - Posterior Probability: 0.00

2\textsuperscript{nd} – half - Posterior Probability: 1.00
Divergent Signals Between Halves

Site index

Support grouping

Second half
Area of Interest

Oppose grouping

Maps within V4 loop of gp120

NIJ grant: 2011-DN-BX-K534
doi:10.1371/journal.pmed.0020337
http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020337
NGS in HIV forensics
Development of the ‘pathogen toolkit’

Long-range PCR

Clone analysis: EcoRI

Large insert cloning

NIJ grant: 2011-DN-BX-K534
NGS in HIV forensics
Development of the ‘pathogen toolkit’

Molecular clone 01
Fragment, add forward adaptor with MC01, reverse adaptor with CS01

Molecular clone 20
Fragment, add forward adaptor with MC20, reverse adaptor with CS01

Molecular clone 01
Fragment, add forward adaptor with MC01, reverse adaptor with CS01

Molecular clone 20
Fragment, add forward adaptor with MC20, reverse adaptor with CS02

Case sample 01

Case sample 02

so forth

Pool libraries, then clonally amplify & sequence by NGS technologies

NIJ grant: 2011-DN-BX-K534
Further Cases and Uses

Viology Journal

Research

A HIV-1 heterosexual transmission chain in Guangzhou, China: a molecular epidemiological study
Zhigang Han†1, Tommy WC Leung2*, Jinkou Zhao†3, Ming Wang1, Lirui Fan1, Kai Li4, Xinli Pang4, Zhenli Li5, Wilina WL Lim2 and Huifang Xu*4

Address: Guangzhou Center for Disease Control and Prevention, Guangzhou, PR China; *Viology Division, Public Health Laboratory Services Branch, Center for Health Protection, Hong Kong SAR, Hong Kong, PR China and †Hainan Center for Disease Control and Prevention, Haikou, PR China

E-mail: Zhigang Han - zhiganghan@gzh.gov.cn; Tommy WC Leung - tommycwleung@yahoo.com; Jinkou Zhao - jinkouhao@hotmail.com; Ming Wang - wangming@gzh.gov.cn; Lirui Fan; Kai Li; Xinli Pang; Zhenli Li; Wilina WL Lim; Huifang Xu

Published: 25 September 2009
Received: 23 June 2009
Accepted: 25 September 2009

This article is available from: http://www.viology.com/content/11/1.4

© 2009 Han et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Haemophilia

Haemophilia (2012), 18, 291–299
DOI: 10.1111/j.1365-2516.2011.02620.x

ORIGINAL ARTICLE Transfusion transmitted disease

Molecular evidence of HIV-1 transmission in 20 Korean individuals with haemophilia: phylogenetic analysis of the vif gene

Y.-K. Cho, § Y. Jung, * J.-S. Lee* and B. T. Foley‡

*Department of Microbiology, University of Ulsan College of Medicine, Seoul, South Korea; and ‡HIV Databases, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, USA

Using HIV Transmission Networks to Investigate Community Effects in HIV Prevention Trials

Joel O. Wertheim†1, Sergel L. Kosakovsky Pond2, Susan J. Little2, Victor De Gruttola3

1 Department of Pathology, University of California San Diego, San Diego, California, United States of America; 2 Department of Medicine, University of California San Diego, San Diego, California, United States of America; 3 Department of Biostatistics, Harvard University, Cambridge, Massachusetts, United States of America

Molecular Epidemiology of HIV-1 Transmission in a Cohort of HIV-1 Concordant Heterosexual Couples from Dakar, Senegal

Wim Jennes1*, Jordan K. Kyongo1, Evelyn Vanhommerig1, Makhtar Camara2, Sandra Coppens3, Moussa Saydi4, Souleymane Mboup5, Leo Heyndrickx6, Luc Kestens2

1 Laboratory of Immunology, Department of Biomolecular Sciences, Institute of Tropical Medicine, Antwerp, Belgium; 2 Laboratory of Immunology, Department of Biophysics, University of Wageningen, Wageningen, The Netherlands; 3 Laboratory of Virology, Department of Biomolecular Sciences, Institute of Tropical Medicine, Antwerp, Belgium; 4 Department of Infectious Diseases, Centre Hospitalier Universitaire Bruxelles, Brussels, Belgium; 5 Laboratory of Virology, Department of Biomolecular Sciences, Institute of Tropical Medicine, Antwerp, Belgium; 6 Department of Public Health, Université Libre de Bruxelles, Brussels, Belgium
MRSA Transmission on a Neonatal Intensive Care Unit: Epidemiological and Genome-Based Phylogenetic Analyses

Ulrich Nübel1*, Matthias Nachtebel2,3,4, Gerhard Falkenhorst2, Justus Benzler2, Jochen Hecht5,6, Michael Kube5, Felix Bröcker5, Karin Moelling5,7, Christoph Bührer8, Petra Gastmeier9, Brar Piening9, Michael Behnke9, Manuel Dehnert2, Franziska Layer1, Wolfgang Witte1, Tim Eckmanns2

1 Department of Infectious Diseases, Unit of Nosocomial Infections, Robert Koch Institute, Wernigerode, Germany, 2 Department for Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany, 3 Post Graduate Training in Applied Epidemiology, Robert Koch Institute, Berlin, Germany, 4 European Programme for Intervention Epidemiology Training (EPIET), European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden, 5 Max Planck Institute for Molecular Genetics, Berlin, Germany, 6 Berlin-Brandenburg Center for Regenerative Therapies, Charité University Medical Center, Berlin, Germany, 7 Institute of Medical Virology, University of Zürich, Zürich, Switzerland, 8 Department of Neonatology, Charité University Medical Center, Berlin, Germany, 9 Institute of Hygiene and Environmental Medicine, Charité University Medical Center, Berlin, Germany
Acknowledgements

BCM-HGSC
Diane I. Scaduto – UTMDA
Xiao-Mei Liu
Wade C. Haaland *
Donna Muzny
Eric Boerwinkle
Richard A. Gibbs

U of Texas
Jeremy M. Brown
Derrick J. Zwickl
David Hillis

U of Michigan
David P. Mindell
Roger G. Ptak