MICROBIAL FORENSIC BIOCRIMES AND HIV

Bruce Budowle[1], Rowan Campbell[2], Arthur Eisenberg[2], Mark Wilson[1], and Ranajit Chakraborty[3]
[1]*FBI Laboratory, Quantico, VA*
[2]*University of North Texas health Science Center, Ft. Worth, TX*
[3]*Center for Genome Information, Dept. of Environmental Health, University of Cincinnati, Cincinnati, OH*

Introduction

Infectious diseases account for 29 out of the 96 major causes of human mortality and morbidity and about 25% of global deaths (i.e., 14 million) per year (64). Advances in molecular biology and microbiology have and will provide tools for combating a number of these diseases. Unfortunately, these technological boons to modern medicine can also be utilized to generate, as well as create, new devastating biologic weapons that can be used nefariously in acts of terrorism and/or to commit biocrimes. Indeed, today the capability to manipulate and disseminate pathogens has enhanced to a point where there is grave concern about the potential use of microbiological organisms to attack and cause serious harm to humans, animals, and plants. In addition to inflicting harm, bioterrorists and biocriminals use microorganisms or toxins as weapons to invoke fear and/or affect economic well-being. Significant effort must be put forth to identify individuals who use such weapons and to prevent or minimize their ability to execute acts of bioterrorism or biocrime. While perpetrators of both bioterrorism and biocrimes use pathogens to inflict harm or create fear, we distinguish the latter as being similar to traditional crimes which usually are directed towards harming specific individuals except that the weapon is biological in nature. However, the use of a bioweapon in a criminal case should be taken as seriously as that of a bioterrorist attack.

The recent bioterrorism attacks with anthrax laden letters and the need to be able to identify perpetrators of bioterrorism and biocrimes have prompted the federal government to develop a new forensic science discipline known as Microbial Forensics. Microbial Forensics is a scientific discipline for analyzing evidence from acts of bioterrorism, biocrime, or the inadvertent release of a microorganism or toxin. Forensic microbiological investigations are essentially the same as any other forensic investigation. They involve crime scene(s) investigation, chain of custody practices, evidence collection and handling, evidence shipping, analysis of evidence, interpretation of results, and court presentation. Molecular analyses of microbial agents from crime scenes are somewhat analogous to those used in the human forensic DNA arena. However, one should be cautious in attributing the meaning of genetic “fingerprinting” to a pathogenic agent, as is often achieved with human DNA analyses. Unique genetic identification of a microorganism may never be possible, because of the clonal nature of many microorganisms, less than optimal population and phylogenetic data, and in some instances limited historical and epidemiological information. Nonetheless, by understanding the limits of the analyses and data, interpretations of genetic evidence can be made qualitatively and/or quantitatively.

When genetic profiles from microbial evidentiary and reference samples are compared, a variety of issues may arise. For example: What might be deduced about the source of the evidentiary sample? Are the samples from the same source or lineage? Are the genetic differences too few to conclude the samples are from different sources (or different lineages)? Are these differences sufficient to consider that the samples are from different sources? Is it possible that the two samples have a recent common ancestor? The degree that these questions can be addressed depends on the context of the case. For example, because they are maintained under controlled conditions, laboratory stocks may show less diversity than samples found in nature. Thus, one or two genetic differences between evidentiary and reference samples may be significant if the weapon originated from a laboratory maintained culture. Alternatively, some genes in bacteria (e.g. *E. coli* and *Salmonella*) have elevated mutation rates (~100-fold) due to defects in mismatch repair, presence of mutator genes, etc.(37, 62). Thus, it is expected that even two recently linked isolates may differ at these rapidly evolving sites. Stable genetic elements should be interpreted differently than rapidly evolving sites. Such knowledge as general source (i.e., laboratory, nature), stability of genetic sites or elements, storage conditions, mutagenic treatments, etc. can further refine interpretations. If lineage identity is the goal rather than precise source identification, one or a few

distinguishing genetic markers may provide sufficient evidence to support or exclude certain hypotheses. Regardless, it is anticipated that the majority of microbial forensic genetic evidence will fall into the class characteristic category instead of establishing virtual identity. In some scenarios statistical weight may not be feasible to assess, yet qualitative evaluations may still be made.

To illustrate some of the issues and considerations for the proper interpretation of microbial forensic evidence, let us consider a biocrime scenario. A number of biocrime cases have been well-documented (10, 13, 22, 44, 50, 51, 58), and unfortunately more are likely to occur. Most of the microbial forensic approaches for biocrimes generally apply to bioterrorism cases. Epidemiology plays an important role in this type of forensic investigation and therefore forensics and epidemiology can not be thought of as separate disciplines in the context of a crime. Epidemiologists attempt to determine the distribution, causes, and risk factors of health problems in specified populations. In microbial forensics, the epidemiologic causes are specific disease causing pathogens and their toxins, their modes of transmission, and man-made manipulations used intentionally against human, animal, and/or plant targets. A forensic investigation will attempt to determine the source of the bioweapon in much the same manner as in an epidemiologic investigation. Genetic analysis can assist in elucidating relationships, even if the population data and genetic diversity may not be well-defined. To illustrate, we describe a mock biocrime and some real criminal investigations involving the human immunodeficiency virus type 1 (HIV).

HIV is an excellent example of a pathogen that naturally causes disease and is not generally considered a primary bioweapon, but yet still can be used in crimes. A number of criminal investigations have involved forensic analyses of HIV (13, 45, 50, 51, 58). Because HIV is a rapidly evolving virus (39), it is illustrative of some of the complexities in a genetic investigation that go beyond simple direct matching of nucleic acid sequences or of genetic marker types. It is highly likely that two HIV samples with a recent common origin will differ at a number of nucleotides within the genome. Also, since there are multiple HIV variants within a host, sequences from the donor may not exactly match those found in a direct transmission recipient . Yet, tools, such as phylogenetic analysis, at the disposal of the microbial forensic analyst (and epidemiologist) can be useful in supporting or refuting relationships of those isolates that have an alleged recent common ancestry compared with those isolates that do not. The bioinformatic tools already exist to assist in such genetic analyses (for example see ref 39).

To begin, we give a brief introduction on the basic structure and genetics of HIV. Next, a short description is provided on phylogenetics and how it is used to assess genetic lineage relationships. Then, a mock case is described to emphasize the value of epidemiology and genetic analyses. Lastly, three actual cases are briefly described to demonstrate that the use of reliable genetic tools and epidemiology have been informative in the solution of HIV-related biocrime cases. At the outset, we should also mention that a thorough review of HIV and phylogenetic approaches is not possible here. The reader should refer to cited literature sources for more details.

Basic Biology of HIV

Although not the premise of this paper, a short introduction into the complexity of HIV may be useful. HIV is a RNA containing icosahedral retrovirus with an envelope that has protruding spikes (Figure 1). The RNA molecule is 9 kilobases in length and contains 9 genes that encode 15 different proteins (Figure 2). These proteins include: gp120, gp41, reverse transcriptase (RT), protease (Pro), integrase, tat, rev, vif, vpr, vpu, nef, RNase H, and matrix protein. The *pol* gene encodes the three essential enzymes pro, RT, and integrase. The *env* gene encodes surface glycoproteins, such as gp120 and gp41. The *gag* gene (group antigen) encodes major structural proteins inside the envelope. The regulatory proteins are encoded in the genes *tat* and *rev* (a promoter), and accessory proteins are encoded by *nef*, *vif*, *vpr*, and *vpu* (19, 23, 27).

To infect the target cell (a helper T cell) HIV must inject its genetic material into the host cell cytoplasm. The viral envelope fuses with the host cell membrane by attaching via the CD4 receptor and other co-receptors on helper T cells. The viral envelope proteins gp120 and gp41 facilitate attachment. After

entering the cell, the viral coat is removed, and the RNA genome is released into the host cytoplasm. A double-stranded viral cDNA molecule is synthesized utilizing the HIV genome as a RNA template and the enzyme RT that was housed within the viral capsid. RNase H removes the viral RNA template molecule from the newly synthesized DNA strand so that a complementary DNA strand can be made off of the newly synthesized DNA molecule. Thus, reverse transcription produces the HIV preintegration complex (PIC), which is composed of double-stranded viral cDNA, integrase, matrix protein, vpr, RT, and the high mobility group DNA-binding cellular protein HMGI(Y). The protein vif is essential for replication and appears to stabilize the reverse transcription complex. The newly synthesized DNA enters the nucleus, facilitated by the matrix protein, and integrates (provirus) into the host cell genome mediated by the integrase. The integrase has three functional activities. Its exonuclease activity removes two nucleotides from each 3' end of the viral cDNA duplex. Then, the integrase cleaves the host DNA at a specific site of integration using its double-stranded endonuclease activity. Lastly, the integrase's ligase activity covalently links the proviral DNA into the host genome. Thus, the provirus is maintained within the cell and is passed on to subsequent daughter cells. The provirus subsequently is transcribed into new viral RNA molecules facilitated by the proteins tat, rev, and nef. Tat, a transcriptional transactivator protein, is essential for HIV replication. Rev, a promoter protein, facilitates the export of viral RNA from the nucleus to the cytoplasm. Nef plays a role in optimizing viral replication and stimulating HIV infectivity. The transcripts are translated into viral proteins using the host cell's RNA polymerase. Pro, integrase, RNase H, and RT are always expressed as the gag-pol fusion protein. The pro cleaves the gag and gag-pol polyprotein precursors, so that the specified proteins will be available for virion assembly. New capsids assemble around viral RNA molecules and necessary proteins. Vif may play a role in the assembly of the virions. Each virion contains about 1500 molecules of gag, about 100 gag-pol polyproteins, two copies of the viral RNA genome, and about 100 molecules of vpr. The vpr protein seems to have a role in enabling HIV to infect non-dividing cells via nuclear localization of the PIC. The virus leaves the cell by budding. Gag precursor protein is involved in initiating budding. While, vpu enhances the release of virions from the host cell (19, 23, 27).

One of the major characteristics of HIV is its genetic variability. RNA viruses and retroviruses, in general, fix substitutions approximately a million times faster than their eukaryotic hosts (18, 60). One reason for the high heterogeneity is the inherent error rate of RT. The HIV RT lacks 3'- to 5'-exonuclease proofreading capability (48, 54, 62). Thus, the misincorporation rate is high during DNA synthesis and as many as five- to-ten errors per genome per cycle of replication occur (49, 55, 63). This lack of proofreading activity is one of the major mechanisms for generating genetic variation within HIV populations within a host or within a population (49, 55, 63). The high variability of HIV can also be attributed to the large number of generations per unit time and the recombinogenic nature of RT. HIV has a mutation rate that can be at least two orders of magnitude greater than mutations found in the eukaryotic germ line DNA. Some of the highest mutation rates are observed at sites within the env gene region for the gp120 glycoprotein, the RT gene, and the gag gene region (5, 8, 24, 26, 32, 34, 35, 36, 41, 44, 65, 66).

The HIV genome is composed of two copies of single stranded-RNA. Both RNA strands are utilized as templates during reverse transcription. Because of this diploid nature, it is possible for RT to jump from one template RNA to the other copy during DNA synthesis. HIV recombines approximately two to three times per genome per replication cycle, which is a high rate of recombination given its relatively small genome (3, 28, 31, 74). HIV can recombine with other related retroviruses. Thus, if two different HIVs infect the same cell, a newly formed virus could contain one RNA molecule from each strain. Notably different genomic combinations can be created in individuals infected by genetically diverse viruses. RNA recombination contributes notably to HIV genetic variability (7, 9, 16, 21, 29, 46, 54, 56, 73).

The genetic variation of HIV within an individual is lowest upon initial infection. Often, only one or a few variants can be found shortly after infection (17, 40, 53). During this early phase, the HIV sequences in the newly infected recipient individual will be more closely related to a subset of variants from the donor than the cadre of source host variants are to each other. As the course of infection continues, viral sequence variation within a host increases dramatically (4, 14, 15, 69). But both isolates will still be more closely

related than other samples with more distant histories. This phenomenon can be exploited to demonstrate whether or not there is a relationship between an alleged HIV donor and an infected individual (13, 22, 45, 51, 58).

Phylogenetics

Phylogenetic analyses have not often been used in forensic analyses, but may prove useful when evaluating potential microbial relationships (13, 22, 45, 51, 58). Phylogenetics is a systematic approach for identifying and understanding evolutionary relationships by attempting to infer a process from a pattern(s) (25, 33, 65). Based on evolutionary theory, similarity among genes, individuals, populations, and/or species is attributed to common descent or a relationship with a common ancestor (although convergent evolution may be considered at times, possibly when only one trait is evaluated). For microbial forensics, we will employ a subset of phylogenetics known as molecular phylogenetics. This methodology uses changes in DNA markers or sequences to reconstruct an evolutionary history of the microorganisms of interest to distinguish similar or closely related individuals from other lineages. For example, we would expect that two microbial genomes that diverged recently should have fewer sequence differences than two microbial genomes whose common ancestor is further in the past (47).

The typical manner of presenting these evolutionary relationships among a group of individuals is by displaying a hierarchical branching diagram or phylogenetic tree (25, 33, 67). The tree basically is a graphic representation composed of nodes and branches, where only one node connects two adjacent branches. The nodes represent features that are common to a group of taxonomic units. A taxonomic unit can be data from an individual organism, a population, or a species. The branches display the relationship among the taxonomic units based on history, descent, and/or ancestry. The topology of the tree is its branching pattern. The branches can be scaled so they represent the degree of evolutionary change and/or time of change. Alternatively, the branches can be unscaled, in which case there is no direct relationship with either time or degree of evolutionary change. Depending on the method, the length of a branch can be indicative of the distance between two taxonomic units. In addition, trees can be rooted or unrooted. Rooted trees have a node that separates a common ancestor from all other taxonomic units. Rooted phylogenetic analyses include an outgroup, which is a taxonomic unit closely related to the other taxonomic units but has diverged before the common ancestor of those taxonomic units. An outgroup roots the tree and gives directionality to character evolution. Unrooted trees also depict a network of relationships among the taxonomic units without necessarily identifying a common ancestor or evolutionary path.

Cladistics, where the search is for a strictly bifurcating tree, is one approach for drawing relationships among taxonomic units (25, 33, 67). The foundation of cladistics is, as described above for phylogenetics in general, that individuals belonging to a particular group share a common evolutionary history. Members of a group are more closely related to each other than they are to other groups. Related groups share some features. Shared derived characteristics are known as synapomorphies. Terms, such as monophyletic, paraphyletic and others, are used to describe groupings of taxa. The relationships are described by the taxonomic units residing within a clade. A clade is a group of taxonomic units that share a common ancestor (i.e., a monophyletic pattern). Monophyletic means that the shared characters (e.g., sequences) of the taxonomic units derive from a single common ancestor. The monophyletic taxon is the building block of cladistics. A paraphyletic group contains only a portion of all the descendants of a common ancestor. To illustrate, consider Figure 3. Different population groups are labeled A through G. The overall pattern is monophyletic because it consists of a single ancestral population group (A) together with all of its descendant population groups (B-G). Populations D, E, F, and G are a monophyletic group, where D is the ancestral populations and all its descendants (E, F, and G) are contained within the grouping. Population F, including population G, is another monophyletic grouping. However, population F by itself is considered paraphyletic, because it is not defined simply as a population group and all its descendants, but as a clade (F) minus another clade (G). The same grouping designations for F and G apply to populations B and C. B and C are a monophyletic grouping, but B alone is a paraphyletic pattern. The analytical methods of phylogenetics generally are divided into distance-based and character-based

approaches (25, 33, 67). Character-based methods evaluate a collection of character changes, such as individual substitutions among the variant sequences, to estimate likely ancestral relationships. Distance-based methods calculate the distances between all pairs of sequences and then build trees based on those distances. Neighbor Joining is a distance-based method that has been used to assess relationships among isolates of HIV. Examples of character-based methods are maximum parsimony (MP) and maximum likelihood (ML). MP selects different tree topologies by identifying the one that involves the shortest evolutionary pathway. That is the pathway that requires the fewest number of nucleotide changes from an ancestral sequence to all of the current sequences that have been compared. ML is more of a statistical inference method. It attempts to locate a tree that best explains the data given a defined stochastic model of molecular evolution. With ML, a probabilistic model of evolutionary change between characters is utilized to construct trees. Most phylogenetic methods will work well for assessing relationships among HIV infected samples. For that matter, not every tree branch needs to be absolutely correct. In fact, a phylogeny is a hypothesis built on assumptions, so it may not be possible to assert that every branch is entirely correct. Yet for forensics, it is more important that the sequences of interest are clustered with the correct subtype(s); reasonable similarity among isolates may be sufficient information for a particular case. Establishing similarity will depend on the samples and time difference from common ancestry, intermediate hosts (if any), the genetic region(s) analyzed, base composition and substitution/indel rates between sites, etc. (38). ML has been used in forensics for assessing a hepatitis C outbreak (22) and a HIV infection by intentional injection (45). This method can provide a statistical framework using a mismatch distribution for assessing the likelihood of the given evidence under the hypotheses of a particular sample(s) belonging to one group of reference samples or not. Even though there is substantial heterogeneity within and among HIV isolates, phylogenetic analyses have identified three main HIV groups, M (major), O (outlier), and N (new) (1, 2, 21, 43, 52, 57, 61, 71, 72). Group M sequences are those ones responsible for the current pandemic. They are the most prevalent and are believed to have originated from a single chimpanzee-to-human transmission event (42). The HIV M group is further subdivided into subgroups A-K. The HIV groups O and N are far less prevalent than group M and also are believed to have originated from a chimpanzee-to-human transmission, but as separate events from each other and from that of group M (20, 30, 42, 48, 57) . At least three-to-five subtypes of O strains have been identified, but nomenclature for group O has not been formalized (59, 70). Group N appears to be the result of a recombinant event between group M and O (42), which may call into question the chimpanzee origin of group N.

HIV - A Potential Weapon in Criminal Cases

Before proceeding with examples of microbial forensic analyses involving HIV, we would like to evaluate the potential for HIV to be used as a weapon. The Scientific Working Group on Microbial Genetics and Forensics (SWGMGF) has proffered the following criteria for assessing a pathogen's potential as a bioweapon. These are: accessibility, culturability, stability, capability of dissemination, environmental tolerance, infectivity, host susceptibility, existence of therapies and control mechanisms, epidimicity, and potential harm to the perpetrator.

Using these criteria, HIV does not appear to be a good tool for biowarfare. Particularly, it is unstable in the environment, and it is not readily transmitted without intimate contact or through fresh blood-blood transfer. On the other hand, HIV does persist in the host for long time periods (latency period can be greater than 10 years); thus, increasing the odds of transmission. Also, the virus is accessible because a large number of people world-wide harbor HIV. These traits make HIV better suited for use in a biocrime than in biowarfare, but terrorists could still employ HIV. While HIV has not been used by a terrorist group to impact public security, it has been used in the committing of some biocrimes (i.e., personal criminal attacks). In fact, a number of states and other countries have passed laws against knowingly transferring HIV by rape or by not informing sexual partner(s) of being HIV positive, by health care transmission through improper unsterile conditions, and by criminal threats with blood or body fluids (e.g., using a needle point as a weapon, spitting, or deliberately injecting someone with HIV) (http://www.ithaca.edu/faculty/shevory/vita/HIV.htm).

Mock Biocrime

To illustrate how a microbial forensic investigation may ensue, consider the following scenario. Ten seemingly unrelated individuals become infected with HIV. All have had surgery at the same hospital in the past year within two weeks of each other. All received a blood transfusion and all have the same blood type. There were no complications during any of the surgeries. Such a scenario naturally raises issues such as the possibility of the cases being related, the possible modes of transmission, and possible suspects if it is deemed an intentional or irresponsible act. It would seem obvious that the hospital is the focal point. But, an epidemiologic approach for disease etiology should be carried out to rule out alternate hypotheses (such as the endemicity of the specific strain or substrain of HIV in the locality). Where possible, risk factors should be assessed. In this case, the patients claim they have no known association with each other. Their life styles are investigated (by questioning the individuals), and it appears they all were at low risk for contracting HIV. All patients were tested for HIV prior to surgery, and all were HIV negative. Thus, it seems unlikely that the patient's activities prior to entering the hospital contributed to any of the HIV infections. After leaving the hospital, their activities were unrelated to one another and not substantially different from that before entering the hospital. Thus, the epidemiologic information points to some event in the hospital, which originally seemed the likely focal point.

Documentation showed that all units of blood used for transfusions were tested for HIV and were negative. This suggests that one of the health care workers may be infected and transferred HIV to the patients during invasive surgery. The work records of all health care providers on staff were reviewed, and eighteen people were identified who had access to any of the patients prior to surgery, during surgery and in post operation recovery. At his/her last physical, each health care worker tested negative for HIV. All willingly provided another blood sample, and all still tested HIV negative. Thus, the health care workers that were in contact with the patients do not seem to be the source of the HIV.

Viral isolates from the patients were obtained and cloned to isolate individual variants. A subset of the variants from each patient were amplified by PCR and sequenced at the RT, gp120 and gp41 regions. In addition, control samples (i.e., HIV positive individuals from the same community as the patients) were analyzed. Figure 4 shows the ML phylogenetic tree. The patients cluster and have short branch lengths compared with the control isolates. Although the donor of the HIV infections has not been identified, the phylogenetic analysis suggests a common source for the HIV infections.

It seems that the source of the HIV infection would remain a mystery, except that the epidemiologic investigation turned to reviewing hospital records for HIV patients who entered the hospital within two weeks prior to the patients' surgeries. Only three HIV positive patients were identified. All three willingly provided blood samples for phylogenetic analysis. The variant strains from one of the three HIV positive patients clustered with the ten infected patients; the other two sample isolates were quite distant from the cluster. However, this one patient had no contact with the other ten patients and does not appear to have been a direct transmitter of the HIV. Yet, it is plausible that his blood may be the source (or he is linked to a recent intermediary).

The investigation identifies a medical technologist who had drawn the blood from the HIV positive source patient. The technologist was also responsible for storage of the blood units in the hospital. The authorities interrogate the technologist, and he confesses to tainting the blood units with HIV positive blood.

This is a fictitious case, and we have no reason to believe such has ever happened or if it is even possible to perpetrate such a crime. It may not be possible to breach a blood packet without leaving obvious marks or damaging the container. The purpose of this example was to illustrate a microbial forensic investigation and how it may rely on epidemiologic, molecular biology, and phylogenetic analyses. The epidemiologic investigation relied on the veracity of the patients and the willingness of individuals to provide blood samples. All the patients and possible blood donors may not be compliant in an actual case. Furthermore, the hospital may not be willing to disclose records on other patients. However, partial information may still be useful in supporting or ruling out certain scenarios. Even in this mock case, the true source of the

sample was not initially available for analysis. Yet, a common source for the HIV was supported by the phylogenetic analysis of specific gene regions in the viral variants. Note that no quantitative weight of the genetic relationship was needed to support the hypothesis that all the patients' HIV were related. A qualitative assessment was sufficient.

Examples of Documented HIV Biocrimes

We suggested above that HIV was more likely to be used in biocrimes than in acts of bioterrorism. It is conceivable that the above mock case could be considered a terrorist act and not a personal crime. We cannot predict the motivation or logic (or lack thereof) of a terrorist, thus any potential disease causing agent may be exploited, even if it is not a top category pathogen. Historically, though, criminal cases involving HIV have not been terrorist acts. The three HIV biocrime scenarios most likely to occur are knowingly transferring HIV by rape or by not informing sexual partner(s) of being HIV positive, by health care transmission through improper unsterile conditions, and by criminal threats with blood tainted items. All three scenarios have occurred and a brief example of each is discussed below.

Transmission Via Invasive Oral Surgery

One of the most well-known HIV forensic cases attempted to determine the cause and source of HIV infection in six individuals in Florida. The patients were allegedly infected by David J. Acer, a dentist, who had treated all six individuals. The six people in question became infected with HIV in the late 1980's and did not indicate life styles or practices that would put them at high risk for exposure to HIV (13). The epidemiologic investigation suggested that a more plausible explanation was that HIV transmission occurred during invasive dental care from their dentist who had AIDS. The dentist was first identified as HIV positive in 1986 (11, 12) and continued to practice. The only data implicating the dentist as the source of the HIV was that he performed invasive oral surgery on the patients. To provide further insight into whether or not the dentist transmitted HIV to his patients, a phylogenetic analysis was performed. Sequence data from the HIVs from lymphocytes from each of the patients, the dentist, a local control group and an outgroup were aligned and compared. The analysis (e.g., sequencing of the gp120 region) showed that the HIV sequences from five of the patients were closely related (although not exactly the same) to those from the dentist and were distinct from viruses obtained from control patients living in the same geographic area as the dental practice. These data strongly supported the hypothesis that the HIVs from the patients were closely related to the HIV from the dentist and that they had contracted the virus from the dentist (11, 12, 13, 50, 51, 68). One patient had notably different viral variants from that of the dentist and therefore, the dentist was eliminated as being the source of that patient's infection (68).

The Florida dentist case is not without controversy (6). The epidemiologic investigation was questioned. Even a 60 Minutes broadcast was aired about the patients, and possible contacts other than Dr. Acer were proffered. One of the female patients was alleged to have lied about sexual relationships and had a disease that could have been sexually transmitted. Another female patient had an affair in the 1970s and her partner was never tested for HIV (even though the AIDS epidemic began after the time of the affair). Such controversies, either legitimate or not, are likely to play out in the court room, but also point to the need to be cautious with epidemiologic data. Thus, the use of reliable scientific methods, such as phylogenetic analysis, will be invaluable in evaluating the circumstances of a particular case, especially when other data may be more difficult to verify.

HIV Sexual Transmission Case

In 1997-1998 in upstate New York, an investigation was carried out involving Nushawn Williams who was accused of being, via sexual transmission, the source of HIV infection in 13 females (58). He had 42 primary female contacts, and 13 tested positive for HIV. Blood samples were obtained for phylogenetic analysis (of the *env* gene and a region of *gag* gene) from 10 of the 13 women, one HIV-infected secondary contact (the only HIV positive secondary partner of the women out of 50 typed), and two HIV-infected individuals from the local community. Mr. Williams did not provide a blood sample. The phylogenetic

analyses demonstrated that the sequences from the 10 HIV positive primary female contacts clustered. The HIV sequences from the reference strains, the secondary contact, and the local community control subjects isolates were excluded as being related to the 10 primary female contacts. The analysis of the HIV strains in concert with the epidemiologic information (i.e., known modes of transmission of HIV and Mr. Williams contacts) supported that Williams was the source even though his blood was not analyzed.

Infection Via Injection

In 1994, gastroenterologist Dr. Richard J. Schmidt was accused of preparing a mixture of blood from two of his patients, one infected with HIV and one infected with hepatitis C. He then allegedly infected his girlfriend (with whom he was having an affair) by injection under the guise of providing her with a vitamin B shot. After being identified as HIV positive in early 1995, a forensic/epidemiologic investigation began. The victim reported a sexual history of contact with seven men including the doctor. All provided blood samples and tested HIV negative. The victim was a nurse, but had no documented needle sticks. She did report splashing of saliva onto her skin by a HIV infected patient in the mid 1980's, but subsequently tested negative for HIV. In fact, the victim was a blood donor and tested HIV negative through early 1994. Investigators then focused on the doctor. A missing patient log book was found, and the last entry was a blood draw from a HIV positive patient on August 1994. The investigators' hypothesis became that this HIV positive patient was the source of the HIV found in the victim. Blood was collected from the HIV positive patient, the victim, and control samples (i.e., HIV positive individuals in the vicinity where the patient and victim reside). ML phylogenetic analyses were carried out on sequences of regions of gp120 and the RT genes. Clustered sequences from the patient and victim supported the hypothesis that their HIV variants were related (45). Dr. Schmidt was found guilty of second degree attempted murder and is serving a 50 year sentence. The admissibility of the phylogenetic analysis was challenged on the basis that there was no precedence for establishing similarities between the viral infections in different individuals. The appeal was rejected by the Louisiana State Supreme Court in 2000 (Personal communication, B. Korber LLNL; S Sinha, Reliagene).

Metzger, et al. (45), who performed the genetic analyses in this case, exercised caution and did not exceed the bounds of the data analysis when interpreting the results. There are unknown factors that can impact on the analysis, such as the within-host sample size, the stage of disease, drug therapies, and drug effectiveness. In addition, a representative control population was based on limited sampling and a willingness to participate. The genetic data could not demonstrate the route of transmission. A patient-to-victim route was equally likely as a victim-to-patient route. The epidemiology and probable contacts helped clarify the transmission mode. Had the sample from the victim been collected soon after infection, a phylogenetic analysis may have been able to shed light on the direction of transmission. A paraphyletic relationship of the victim's viral sequences with respect to the viral sequences of the patient sample may have supported a patient-to-victim transmission direction. The paraphyletic relationship is lost over time due to lineage extinction. Additionally, the data could not determine if additional individuals were involved between the transmission line of the patient and victim.

Conclusion

As with any forensic analysis, analytical results need not be absolute concerning source attribution to be probative. The examples above demonstrate that reliable genetic tools exist that can assist in establishing or refuting common histories of microorganisms, even for highly variable retroviruses. Other microorganisms with DNA genomes may be analyzed at more stable sites and the general approach could be the same. In some microbial forensic cases sufficient data may be available to provide a quantitative statement regarding the relationship of evidence and reference samples, and in other cases only a qualitative statement may be possible. Both approaches are acceptable provided that the assumptions and limits are defined. We stress that epidemiology should not be thought of as a separate and distinct discipline from microbial forensics. Instead, beyond health and public safety considerations, epidemiology is a foundational part of microbial forensics. Epidemiologic considerations form an integral component of

selection criteria of microbes on which database composition is based and in addressing QA/QC issues.

Lastly, while the goal of a microbial forensic analysis is to determine the microbial source of the pathogen, genetic data may never achieve the level of individualization. Microorganisms mostly have an asexual mode of reproduction. Therefore, a number of isolates may have the same genetic profile and yet not be the most recent source of the bioweapon. We should not lose sight of the role traditional forensic analyses, such as human DNA analysis, fingerprint analyses, analytical chemistry, elemental analysis, tool marks and other techniques will play for attribution of a bioterrorist event or biocrime evidence. In some cases, these traditional approaches may be more informative than the genetic analyses of the microbial genome.

Acknowledgment

This is publication number 03-19 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

1. Ayouba A, Souquieres S, Njinku B, Martin PM, Muller-Trutwin MC, Roques P, Barre-Sinoussi F, Mauclere P, Simon F, Nerrienet E (2000) HIV-1 group N among HIV-1-seropositive individuals in Cameroon. AIDS 14:2623-2625.

2. Bibollet-Ruche F, Loussert-Ajaka I, Simon F, Mboup S, Ngole EM, Saman E, Delaporte E, Peeters M (1998) Genetic characterization of accessory genes from human immunodeficiency virus type 1 group O strains. AIDS Res. Hum. Retroviruses 14:951-961.

3. Bowman RR, Hu WS, Pathak VK (1998) Relative rates of retroviral reverse transcriptase template switching during RNA- and DNA-dependent DNA synthesis. J. Virol. 72(6):5198-5206.

4. Brown AJ (1991) Sequence variability in human immunodeficiency viruses: pattern and process in viral evolution. AIDS 5, Suppl 2:S35-S42.

5. Brown AJ, Lobidel D, Wade CM, Rebus S, Phillips AN, Brettle RP, France AJ, Leen CS, McMenamin J, McMillan A, Maw RD, Mulcahy F, Robertson JR, Sankar KN, Scott G, Wyld R, Peutherer JF (1997) The molecular epidemiology of human immunodeficiency virus type 1 in six cities in Britain and Ireland. Virology 235(1):166-177.

6. Brown D (1996) The 1990 Florida dental investigation: theory and fact. Ann. Intern. Med. 124:255-256.

7. Burns CC, Gleason LM, Mozaffarian A, Giachetti C, Carr JK, Overbaugh J (2002) Sequence variability of the integrase protein from a diverse collection of HIV type 1 isolates representing several subtypes. AIDS Res Hum Retroviruses 18(14):1031-1041.

8. Carneiro M, Yu XF, Lyles C, Templeton A, Weisstein AE, Safaeian M, Farzadegan H, Vlahov D, Markham RB (1999) The effect of drug-injection behavior on genetic evolution of HIV-1. J. Infect. Dis. 180(4):1025-1032.

9. Carr JK, Salminen MO, Albert J, Sanders-Buell E, Gotte D, Birx DL, McCutchan FE (1998) Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 247:22-31.

10. Carus WS (2002) Bioterrorism and Biocrimes: The Illicit Use of Biological Agents Since 1900 Fredonia Books, Netherlands, pp. 1-209.

11. CDC (1990) Possible transmission of human immunodeficiency virus to a patient during an invasive dental procedure. MMWR 39:489–493.

12. CDC (1991) Update: transmission of HIV infection during an invasive dental procedure---Florida. MMWR 40:21–27.

13. Ciesielski C, Marianos D, Ou CY, Dumbaugh R, Witte J, Berkelman R, Gooch B, Myers G, Luo CC, Schochetman G, et al (1992) Transmission of human immunodeficiency virus in a dental practice. Ann. Intern. Med. 116:798-805.

14. Coffin JM (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267:483-489.

15. Coffin JM (1996) HIV viral dynamics. AIDS 10, Suppl 3:S75-S84.

16. Costa LJ, Munerato P, Diaz RS, Tanuri A (2000) Generation of intersubtype human immunodeficiency

virus type 1 recombinants in env gene in vitro: influences in the biological behavior and in the establishment of productive infections. Virology 268(2):440-451.

17. Diaz RS, Zhang L, Busch MP, Mosely JW, Mayer A (1997) Divergence of HIV-1 quasispecies in an epidemiologic cluster. AIDS 11:415-422.

18. Drake JW, Charlesworth B, Charlesworth D, and Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667-1686.

19. Flint SJ, Enquist LW, Krug RM, Racaniello, Skalka AM (2000) In Virology: Molecular Biology, Pathogenesis, and Control, ASM Press, Washington, D.C., pp. 762-763.

20. Gao F, Robertson DL, Carruthers CD, Li Y, Bailes E, Kostrikis LG, Salminen MO, Bibollet-Ruche F, Peeters M, Ho DD, Shaw GM, Sharp PM, Hahn BH (1998) An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). J. Virol. 72(12):10234-10241.

21. Gao F, Robertson DL, Carruthers CD, Morrison SG, Jian B, Chen Y, Barre-Sinoussi F, Girard M, Srinivasan A, Abimiku AG, Shaw GM, Sharp PM, Hahn BH (1998) A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type. J. Virol. 72:5680-5698.

22. Gonzalez-Candelas F, Bracho MA, Moya A (2003) Molecular epidemiology and forensic genetics: application to a hepatitis C virus transmission event at a hemodialysis unit. J. Infect. Dis. 187(3):352-358.

23. Greene WC, Peterlin BM (2003) Molecular insight into HIV Biology. Natural Science of HIV, HIV InSite Knowledge Base Chapter at http://hivinsite.ucsf.edu/

24.. Hernandez-Aguado I, Avino MJ, Perez-Hoyos S, Gonzalez-Aracil J, Ruiz-Perez I, Torrella A, Garcia de la Hera M, Belda F, Fernandez E, Santos C, Trullen J, Fenosa A (1999) Human immunodeficiency virus (HIV) infection in parenteral drug users: evolution of the epidemic over 10 years. Valencian Epidemiology and Prevention of HIV Disease Study Group. Int. J. Epidemiol. 28(2):335-340.

25. Hillis DM, Allard MW, Miyamoto MM (1993) Analysis of DNA sequence data: phylogenetic inference. Methods in Enzymology 224: 456-487.

26. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373(6510):123-126.

27. Hope JT, Trono D (2000) Structure, expression, and regulation of the HIV genome. Natural Science of HIV, HIV InSite Knowledge Base Chapter at http://hivinsite.ucsf.edu/

28. Hu WS, Rhodes T, Dang Q, Pathak V (2003) Retroviral recombination: review of genetic analyses. Front. Biosci. 8:d143-d155.

29. Hu WS, Temin HM (1990) Retroviral recombination and reverse transcription. Science 250:1227-1233.

30. Janssens W, Laukkanen T, Salminen MO, Carr JK, Van der Auwera G, Heyndrickx L, van der Groen G, McCutchan FE (2000) HIV-1 subtype H near-full length genome reference strains and analysis of subtype-H-containing inter-subtype recombinants. AIDS 14(11):1533-1543.

31. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000) High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol. 74(3):1234-1240.

32. Ji J, Loeb LA (1994) Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 env gene. Virology 199(2):323-330.

33. Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) Cladistics, Second Edition, Oxford University Press, Oxford, pp. 1-228.

34. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T (2000) Timing the ancestor of the HIV-1 pandemic strains. Science 288(5472):1789-1796.

35. Kuiken CL, Cornelissen MT, Zorgdrager F, Hartman S, Gibbs AJ, Goudsmit J (1996) Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution. J. Gen. Virol. 77(4):783-792.

36. Kuiken CL, Lukashov VV, Baan E, Dekker J, Leunissen JA, Goudsmit J (1996) Evidence for limited within-person evolution of the V3 domain of the HIV-1 envelope in the Amsterdam population. AIDS 10(1):31-37.

37. LeClerc JE, Li B, Payne WL, Cebula TA (1996) High mutation frequencies among Escherichia coli and Salmonella pathogens. Science 274:1208-1211.

38. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J (1996) Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc. Natl. Acad. Sci. USA 93:10864-10869.

39. Li WH, Tanimura M, Sharp P (1998) Rates and dates of divergence between AIDS virus nucleotide sequences. Mol. Biol. Evol. 5:313-330.

40. Liu SL, Schacker T, Musey L, Shriner D, McElrath MJ, Corey L, Mullins JI (1997) Divergent patterns of progression to AIDS after infection from the same source: human immunodeficiency type 1 virus evolution and antiviral responses. J. Virol. 71:4284-4295.

41. Lukashov VV, Kuiken CL, Vlahov D, Coutinho RA, Goudsmit J (1996) Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. AIDS Res. Hum. Retroviruses 12(12):1179-1183.

42. Mauclere P (2000) HIV-1 group N in Cameroon and apparent viruses in the chimpanzee. Bull. Soc. Pathol. Exot. 93(3):162.

43. Mauclere P, Loussert-Ajaka I, Damond F, Fagot P, Souquieres S, Monny Lobe M, Mbopi Keou FX, Barre-Sinoussi Saragosti S, Brun-Vezinet F, Simon F (1997) Serological and virological characterization of HIV-1 group O infection in Cameroon. AIDS 11:445-453.

44. McCutchan FE, Viputtigul K, de Souza MS, Carr JK, Markowitz LE, Buapunth P, McNeil JG, Robb ML, Nitayaphan S, Birx DL, Brown AE (2000) Diversity of envelope glycoprotein from human immunodeficiency virus type 1 of recent seroconverters in Thailand. AIDS Res. Hum. Retroviruses 16(8):801-805.

45. Metzker, ML, Mindell DP, Liu X, Ptak RG, Gibbs RA, Hillis DM (2002) Molecular evidence of HIV-1 transmission in a criminal case. Proc. Natl. Acad. Sci. USA 99(22):14292-14297.

46. Montavon C, Bibollet-Ruche F, Robertson D, Koumare B, Mulanga C, Esu-Williams E, Toure C, Mboup S, Saman E, Delaporte E, Peeters M (1999) The identification of a complex A/G/I/J recombinant HIV type 1 virus in various West African countries. AIDS Res. Hum. Retroviruses 15(18):1707-1712.

47. Moya A, Elena FS, Bracho A, Miralles R, Barrio E (2000) The evolution of RNA viruses: a population genetic view. Proc. Natl. Acad. Sci. USA 97:6967-6973.

48. Oelrichs RB, Shrestha IL, Anderson DA, Deacon NJ (2000) The explosive human immunodeficiency virus type 1 epidemic among injecting drug users of Kathmandu, Nepal, is caused by a subtype C virus of restricted genetic diversity. J. Virol. 74(3):1149-1157.

49. O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD (2002) Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. J. Biol. Chem. 277(41):38053-38061.

50. Ou CY, Kwok S, Mitchell SW, Mack DH, Sninsky JJ, Krebs JW, Feorino P, Warfield D, Schochetman G, et al (1988) DNA amplification for direct detection of HIV-1 in DNA of peripheral mononuclear cells. Science 239:295–297.

51. Ou CY, Ciesislski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins BT, Schochetman G, Berkelman RL, Economou AN, et al (1992) Molecular epidemiology of HIV transmission in a dental practice. Science 256:1165-1171.

52. Peeters M, Gueye A, Mboup S, Bibollet-Ruche F, Ekaza E, Mulanga C, Ouedrago R, Gandji T, Mpele P, Dibanga G, Koumare B, Saidou M, Esu-Williams E, Lombart J, Badombena W, Luo N, Vanden Haesevelde M, Delaporte E (1997) Geographic distribution of HIV-1 group O viruses in Africa. AIDS 11:493-498.

53. Pieniazek D, Janini LM, Ramos A, Tanuri A, Schechter M, Peralta JM, Vincente AC, Pieniazek NK, Schochetman G, Rayfield MA (1995) HIV-1 patients may harbor viruses of different phylogenetic subtypes: implications for the evolution of the HIV/AIDS pandemic. Emerg. Inf. Dis. 1:86-88.

54. Piyasirisilp S, McCutchan FE, Carr JK, Sanders-Buell E, Liu W, Chen J, Wagner R, Wolf H, Shao Y, Lai S, Beyrer C, Yu XF (2000) A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. J. Virol. 74(23):11286-11295.

55. Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. Science 242(4882):1168-1171.

56. Quinones-Mateu M, Arts EJ (1999) Recombination in HIV-1: update and implications. AIDS Rev. 1:89-100.

57. Rayfield MA, Sullivan P, Bandea CI, Britvan L, Otten RA, Pau CP, Pieniazek D, Subbarao S, Simon P, Schable CA, Wright AC, Ward J, Schochetman G (1996) HIV-1 group O virus identified for the first time in the United States. Emerg. Infect. Dis. 2(3):209-212.

58. Robbins KE, Weidle PJ, Brown TM, Saekhou AM, Coles B, Holmberg SD, Folks TM, Kalish ML (2002) Molecular analysis in support of an investigation of a cluster of HIV-1-infected women. AIDS Res. Hum. Retroviruses 18(15):1157-1161.

59. Roques P, Robertson DL, Souquiere S, Damond F, Ayouba A, Farfara I, Depienne C, Nerrienet E, Dormont D, Brun-Vezinet F, Simon F, Mauclere P (2002) Phylogenetic analysis of 49 newly derived HIV-1 group O strains: high viral diversity but no group M-like subtype structure. Virology 302(2):259-273.

60. Sala M, Wain-Hobson S (2000) Are RNA viruses adapting or merely changing?. J. Mol. Evol. 51(1):12-20.

61. Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F., and Brun-Vezinet, F (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med. 4 :1032-1037.

62. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of E. coli. Nature 387:703-705.

63. Svarovskaia ES, Cheslock SR, Zhang WH, Hu WS, Pathak VK (2003) Retroviral mutation rates and reverse transcriptase fidelity. Front. Biosci. 2003;8:d117-d134.

64. Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356:983-989.

65. Wang N, Zhu T, Ho DD (1995) Sequence diversity of V1 and V2 domains of gp120 from human immunodeficiency virus type 1: lack of correlation with viral phenotype. J. Virol. 9(4):2708-2715.

66. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, et al (1995)Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373 :117-122.

67. Weston PH (1994) Methods for rooting cladistic trees. In: Models in Phylogeny Reconstruction (Scotland RW, Siebert DJ, Williams DM, eds.), Oxford Uni. Press, Oxford, pp. 125-155.

68. Witte JJ, Wilcox Jr KR (1991) Epidemiologic Notes and Reports Update: Transmission of HIV Infection During Invasive Dental Procedures --- Florida MMWR 40(23);377-381.

69. Wolfs TF, Zwart G, Bakker M, Goudsmit J (1992) HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology 189:103-110.

70. Yamaguchi J, Vallari AS, Swanson P, Bodelle P, Kaptue L, Ngansop C, Zekeng L, Gurtler LG, Devare SG, Brennan CA (2002) Evaluation of HIV type 1 group O isolates: identification of five phylogenetic clusters. AIDS Res. Hum. Retroviruses 18(4):269-282.

71. Yang C, Pieniazek D, Owen SM, Fridlund C, Nkengasong J, Mastro TD, Rayfield MA, Downing R, Biryawaho B, Tanuri A, Zekeng L, van der Groen G, Gao F, Lal RB (1999) Detection of phylogenetically diverse human immunodeficiency virus type 1 groups M and O from plasma by using highly sensitive and specific generic primers. J. Clin. Microbiol. 37(8):2581-2586.

72.. Zekeng L, Gurtler L, Afane Ze E, Sam-Abbenyi A, Mbouni-Essomba G, Mpoudi-Ngolle E, Monny-Lobe M, Tapka JB, Kaptue L (1994) Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results. AIDS 8:1626-1628.

73. Zhu T, Wang N, Carr A, Wolinsky S, Ho DD (1995) Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. J. Virol. 69(2):1324-1327.

74. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, Preston BD, Dougherty JP (2002) Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. Virol. 76(22):11273-11282.

Figure 1. Schematic of the structure of HIV. Note that there are multiple copies of the enzymes contained within the virus (see text).

Figure 2. Schematic of the linear proviral genome of HIV displaying relative position of genes described in text. LTR refers to long terminal repeat which is required for initiation of transcription.

Figure 3. A tree of 7 population groups displaying phylogenetic relationships.

Figure 4. Fictitious ML tree from mock biocrime case. p1- p10 refer to sequences from viral isolates from the ten patients. C1-C9 refer to control samples which are from HIV positive people that reside in the same geographic area as the patients. RF refers to a reference HIV sample. OG refers to an outgroup sample. p1-p10 cluster, and all are distant from all other samples.