Oct 4, 2016

Revisiting the DNA analysis of a criminal HIV transmission

- Shishi Luo, John Doe

Background
In 1998, Richard J Schmidt, a gastroenterologist, was convicted of attempted murder of Janice Trahan using the blood of an HIV positive patient under the guise of a vitamin B shot. The transmission occurred in mid-1994, a time when HIV infection was still considered fatal. Among other evidence, the prosecution presented DNA analysis that linked the HIV found in victim with that of Schmidt's HIV positive patient. This was the first example of phylogenetic analysis being admitted as evidence in a US court of law. Note that in comparison, genetic evidence in the form of DNA fingerprinting is much more common in criminal investigations.

Phylogenetics is a term used to describe methods which use genetic data to determine the evolutionary relationships between a set of organisms ('phylo' for group or tribe). This is in contrast to methods which rely of morphological characteristics to group organisms in related species. The general strategy of phylogenetics is to reconstruct a tree representing the relatedness between individuals in a sample such that the distances in the tree are consistent with genetic similarity. 'Consistent' can be construed in various ways, and different phylogenetic reconstruction algorithms use different criteria to judge consistency between the tree and genetic distances. For example, one family of algorithms tries to construct a tree such that the sum of branch lengths between two leaves (individual sequences) is close to the genetic distance between the two sequences.

In 2002, after all appeals had been exhausted, the scientists who conducted the analysis for the prosecution published their results (Metzker et al, 2002). The data they collected was made available with the publication. It is also possible to replicate much of their analysis using publicly available tools.

Aims
There are two aims in this project. The first is to reproduce the DNA analysis conducted by Metzker et al. In their paper, the scientists used *parsimony* and *minimum evolution* methods to determine that the HIV strains found in the victim were more closely related to the doctor's patient than to a sample of HIV sequences from the local population. We will repeat this analysis using tools available on the Los Alamos National Laboratory's (LANL) online HIV database. We will also apply other phylogenetic models on the same dataset to investigate how robust their findings are.

The second aim is to see whether the same conclusions can be drawn using a simpler approach. Instead of conducting a phylogenetic analysis, we will calculate the Hamming Distance between all the HIV sequences in the Metzker et al dataset. Hamming Distance is a cruder measure of similarity between two DNA sequences because it assumes all differences are equally possible (more sophisticated models assume some types of point mutations are more common than others). We will check whether this measure of genetic similarity is sufficient to determine that the victim and patient's HIV are more closely related than that of the sample from the local population.

Potential pitfalls
It may not be possible to reproduce all of the original analyses. The investigators, for example, used the program PAUP to conduct the phylogenetic analyses. Although the LANL HIV database has an

online tool for generating trees using a similar algorithm, subtle differences in the implementation may lead to different biological conclusions.

If the Hamming Distance analysis returns the same result as the phylogenetic analysis, that demonstrates that a simpler model suffices to prove the link between the patient's HIV DNA and the victim's. However, if the Hamming Distance approach is inconclusive or contradicts the original study by Metzker et al, we do not learn much from our exercise.

Distribution of tasks and timeline
The proposed work consists of the following tasks, to be divided as follows:
- Data collection: Shishi Luo (by Oct 8)
- Reproducing Metzker et al analysis: John Doe (by Oct 15)
- Hamming distance analysis: Shishi Luo and John Doe (by Oct 22)
- Figures and visualization: Shishi Luo (by Oct 29)
- Presentation material: John Doe (draft by Nov 5)
- Write-up: Shishi Luo and John Doe (draft by Nov 12)

References and online resources

Metzker, Michael L., et al. "Molecular evidence of HIV-1 transmission in a criminal case." *Proceedings of the National Academy of Sciences* 99.22 (2002): 14292-14297.

Genome News Network report on the Metzker et al (2002) paper: http://www.genomenewsnetwork.org/articles/01_03/hiv.shtml

Appeals case: State v. Schmidt, 699 So. 2d 448 - La: Court of Appeals, 3rd Circuit 1997

Metzker's data is available at the HIV sequence database at http://www.hiv.lanl.gov