

Disease detectives

A powerful method for deducing microbial relationships has been edging its way into civil and criminal investigations. But courts should proceed with caution.

By Shaoni Bhattacharya

Anaesthetist Juan Maeso led a seemingly respectable life in the coastal Spanish town of Valencia. But he had a secret. Over the course of at least a decade, at two different hospitals, he regularly skimmed morphine from his patients, injecting himself just before using the same needle to administer their doses.

In 2007, Maeso was found guilty of infecting at least 275 people with hepatitis C, four of whom had died from complications related to the disease. He was sentenced to 1,933 years in prison, although he is expected to serve only 20 under Spanish law.

To this day, Maeso protests his innocence, saying that a patient must have infected him with the hepatitis C virus (HCV). But the

scientific evidence, which was published in full only last year¹, overwhelmingly suggests otherwise. In that work, Fernando González-Candelas and his colleagues at the University of Valencia analysed and categorized almost 4,200 viral sequences in an effort to disentangle the path the infection followed, using a process known as phylogenetic forensics.

The method, which marries classic evolutionary-biology practices with modern sequencing technology, is increasingly being used in criminal and civil investigations, and for biodefence. A paper published this month², for example, describes how the technique allowed scientists to trace the likely origin of an anthrax-laced batch of heroin that has been

killing users across Europe since 2009.

But the intersection of this science with the legal system makes many uneasy, says Anne-Mieke Vandamme, an evolutionary geneticist at the University of Leuven in Belgium, who has worked on 19 criminal cases since 2002, mostly for the defence. Unlike DNA evidence, which is routinely used in legal settings around the world, the results of phylogenetic forensics are rarely definitive. “You can never prove guilt,” she says.

And there are social concerns. Many patient advocates feel that tracing the path of infection in civil and criminal cases may further stigmatize diseases such as AIDS. Now, as the field matures thanks to advanced sequencing

ILLUSTRATION BY KATIE SCOTT

and analytical tools, a team of experts led by Vandamme is trying to develop guidelines for best practice both on technical aspects of the work and on presenting the evidence in courts. She hopes, she says, “to make clear to lawyers, judges and prosecution officers the powers and limitations of these methods”.

A common factor

Maeso's misdeeds started to come to light when doctors at Spanish utility companies noticed clusters of HCV among workers. While reviewing the workers' medical records, one doctor, Manuel Beltran, noticed that they had all had minor surgery at the Hospital Casa de Salud in Valencia some months before.

Beltran contacted the local public-health authority, sparking what turned out to be a massive investigation that scoured the records of more than 66,000 patients across two hospitals. Early on, it was clear that Maeso was a common factor in many of the cases. But prosecutors would need more evidence.

This is where phylogenetics came in. Some viruses, such as HCV, HIV and influenza, mutate incredibly quickly. By sequencing virus samples from different individuals — and then comparing tiny differences in their genomes — scientists can trace their evolution and place them on a family tree (see ‘Infectious forensics’). “What we are doing is a virus genealogy,” says Oliver Pybus, who studies evolution and infectious diseases at the University of Oxford, UK.

The process allows scientists to predict how likely it is that two or more infections are closely related and what their relationship is. And as the technology has steadily improved, such information has proved increasingly useful. Prosecutors have used it in cases of intentional infections, such as that of Richard Schmidt, who was in 1998 convicted by a Louisiana court of attempted second-degree murder. He injected his former girlfriend with HIV- and HCV-tainted blood, telling her that he was giving her a vitamin B12 jab. The method was used to help track the source of anthrax spores posted to several US politicians and media outlets in 2001. And it has been used to provide evidence in rape accusations and in investigations of child sexual abuse in which a disease was transmitted years earlier.

But phylogenetic evidence is very different in nature from the DNA matches that juries may be more familiar with, says Vandamme: the latter can often confirm or exclude a suspect's involvement in a crime with extremely high certainty. Phylogenetic analyses can offer supporting evidence — that a virus found in person A is very likely to have come from person B, say — but can never prove direct transmission on their own, she says. In the Maeso case, for example, prosecutors used viral phylogenies to corroborate evidence gained from epidemiological investigations.

González-Candelas and his colleagues

used patterns of changes in a highly variable region of the HCV genome to sort the viruses into clades, or branches of a tree that illustrate their evolutionary relationships. The scientists analysed, on average, 11 such viral sequences per person from 321 people believed to have been infected by Maeso and 42 controls — local HCV-infected patients with no known connection to the case. When printed out, the tree that the researchers developed was 11 metres long.

Using all the data, the team determined for each infected individual a ‘likelihood ratio’ — that is, the probability that the infection was related to Maeso's and others whom Maeso had presumably infected, versus the probability that it had come from a source unrelated to the outbreak. Because there were so many samples and a strong phylogenetic signal, the

“
Just because we can
test these relationships
doesn't mean
that it is always in society's
best interest to do so.

”
likelihood ratios the scientists got were high. Most were higher than 10^5 , and the highest was 6.6×10^{95} , exceptionally strong support for this type of analysis.

The Valencia work was also notable in that it attempted to pinpoint when individuals had contracted the virus, using a ‘molecular clock’ technique. To do this, the researchers sampled the genetic diversity of viruses in each person, and then used the mutation rate of HCV in the outbreak to estimate when they had been infected. Almost two-thirds of the estimated dates of infection lined up with when the patients had visited the Valencia hospitals, adding to the evidence that Maeso was the source.

Presenting such data in court is challenging. González-Candelas and his colleague Andrés Moya had to lecture judges and attorneys for two days to familiarize them with evolutionary terms and concepts before launching into three weeks of scientific testimony.

One of the challenges was differentiating the process from conventional DNA testing in minds of the judges and lawyers. Court officials needed to understand that the analysis is inherently more messy: because HCV mutates so rapidly, the longer a person has an infection, the more viral diversity they are likely to have.

When that person infects another, any of the new variants could be passed on, and not all are necessarily sampled in the forensic process, meaning that a connection could be missed or the strength of the relationship distorted. “There is never a full match between strains

of linked individuals or even within a single individual,” says Vandamme. Even in cases in which the viruses from two or more individuals are clearly related, she says, “there are several possible trees, depending on when the samples are taken and how many variants were passed during transmission”.

In the Maeso case, the probabilities linking him to some of the patients were quite strong. But the method also helped to clear him of blame in 47 suspected cases. Those individuals were therefore not entitled to compensation. “Our analysis worked both ways,” says González-Candelas.

Clear cut

Many scientists see the technique's ability to clear individuals of crimes as its greatest strength. In May 2004, five Bulgarian nurses and a Palestinian doctor were sentenced to death for allegedly infecting 426 children with HIV at the al-Fateh Hospital in Benghazi, Libya (see *Nature* 430, 277; 2004). The ‘Benghazi Six’ had been detained and reportedly tortured since 1999.

A phylogenetic analysis had suggested that the particular strain of HIV involved had been circulating years before the arrival of the foreign health workers. *Nature* published the results online just before a retrial in 2006 (ref. 3), and although they did not sway the court from the death penalty at the time, the findings did seem to change diplomatic relations “quite considerably”, says Pybus, who was part of the research team. In 2007, the sentences were commuted to life imprisonment, and the health workers were extradited to Bulgaria, where they were pardoned by the Bulgarian president.

The field has developed since these watershed cases. In 2010, evolutionary biologist David Hillis of the University of Texas at Austin and his colleagues described methods that, for the first time, gave supportive evidence on the direction of viral transmission⁴.

To do this, investigators look closely at the populations of viruses in infected individuals. Because one person harbours many variants, only a subset is transmitted when they infect someone new. Once transmitted, this subset will multiply in number and continue to evolve rapidly. As a result, some viruses in the source may seem to be more closely related to the viruses in a recipient than to other viruses in the source, Hillis explains. Identifying these relationships can help to support a hypothesis of who infected whom.

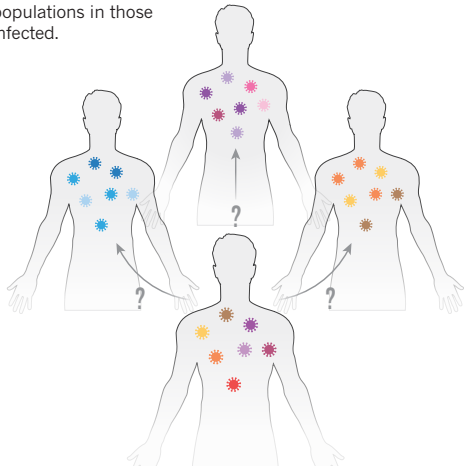
New sequencing technologies are also increasing the power of what phylogenetics can do. “The more you sample, the better — the more you can fill in the gaps,” says Andrew Rambaut of the University of Edinburgh, UK, who worked with Pybus on the Benghazi case.

Rapid, automated sequencing can give a huge amount of information, says Bruce Budowle, who worked as a scientist for the

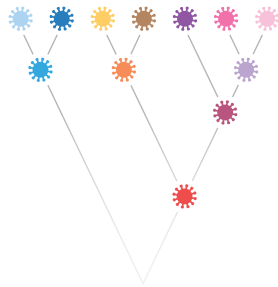
Infectious forensics

Phylogenetics offers a way to establish relationships between microbes infecting several individuals and can be used as corroborating evidence when someone is suspected of infecting others with a disease.

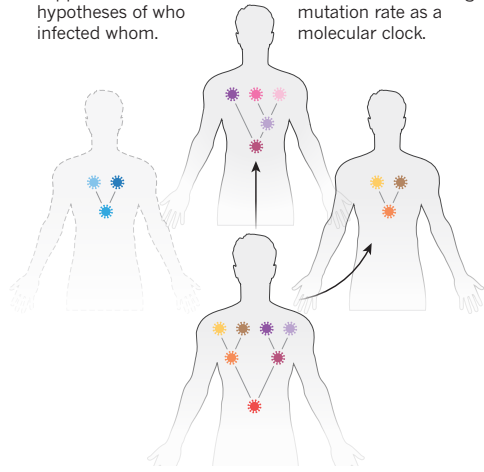
1 Pathogen genomes can mutate quickly, creating diverse microbial populations in those infected.



2 By sequencing highly variable regions of pathogen genomes scientists can build a phylogenetic tree that suggests how the microbes are related.



3 The relatedness of the viral populations can support or rule out hypotheses of who infected whom.



4 Pathogen diversity can also be used to corroborate time of infection using the mutation rate as a molecular clock.

US Federal Bureau of Investigation (FBI) for 26 years, and is now director of the Institute of Applied Genetics at the University of North Texas Health Science Center in Fort Worth.

But there is a catch. The masses of data generated have to be processed in a way that is useful for forensic purposes, he says: if software or methods for developing the phylogenies are not properly validated, findings could be challenged in court. Many useful applications developed in academia may not be subjected to such validation because it is not a priority until the methods are needed for forensics work. “We often get so enamoured with our science, and then something comes up and you have to use it,” Budowle says.

Budowle and his colleagues were in exactly this situation during the 2001 anthrax attacks. To piece together the bacterial phylogenies, they had to use an unvalidated method developed by an academic microbiologist — Paul Keim at Northern Arizona University in Flagstaff. “It gave us guidance on what may have occurred, and pointed to a laboratory strain rather than one found in nature,” says Budowle.

This helped investigators to track the microbe back to a laboratory strain called Ames. A variant of this strain was later linked to Bruce Ivins, a microbiologist at the US Army Medical Research Institute of Infectious Diseases at Fort Detrick in Maryland. It is impossible to say how important the phylogenetic data would have been because the case never went to court. After the FBI began to investigate Ivins in 2008, he committed suicide (see *Nature* 454, 672; 2008).

The high stakes involved in many cases using phylogenetics has led to other concerns — notably, that the technique might contribute to the stigmatization of people infected with HIV, or to the criminalization of HIV transmission. In several countries, people have been charged and prosecuted for murder,

attempted murder or bodily harm for unwittingly transmitting the virus to a sexual partner or not disclosing that they have it — even if it was not transmitted. Some researchers think this dissuades people from coming forward for testing.

For this reason, some in the phylogenetics field have stopped working on criminal cases altogether or are extremely selective about the cases they take on. Andrew Leigh Brown, who studies HIV evolution at the University of Edinburgh, assisted in the first-ever investigations using phylogenetics for forensics in the early 1990s. But he no longer works on such cases.

Leigh Brown contributed to a policy document, published by the Joint United Nations Programme on HIV/AIDS last May, calling for an end to prosecution for HIV transmission other than in clearly intentional cases. Where intent is apparent, phylogenetic forensics should be used carefully and with other supporting evidence, the document advises: the burden of proof must be high.

Promise and pitfalls

Vandamme laments the lack of guidance for researchers in phylogenetic forensics. She hopes the guidelines that she and other concerned specialists are currently drafting will help scientists to avoid misinterpretations. In addition to providing tools for presenting findings in court, Vandamme hopes to reach a consensus on technical issues such as how to find a control population and which genetic regions of a virus should be assessed. “This will help the increasing number of phylogenetic experts that are called by court to provide their expertise in a forensic context,” she says.

Moving forward, scientists say that they will continue to carefully pick which cases they agree to get involved in. “Just because we can test these relationships doesn’t mean that it is always in society’s best interest to do so,”

says Hillis. “My own choice is to work on such analyses only when they are used to test a clear crime that goes beyond accidental viral transmission, such as rape or attempted murder.”

Although the Valencia case is several years old, publication of the data has renewed discussion about phylogenetic forensics, its potential uses and its pitfalls, not just in legal proceedings, but also in biodefence. To that end, González-Candelas was invited to speak at a meeting in Zagreb, Croatia, last October to hammer out the main challenges that the field faces.

The workshop, hosted by the US National Academy of Sciences and the UK Royal Society, among others, has not yet published its findings. But Budowle says that there is a major conflict in the field over access to data, with members of the biosecurity and intelligence communities wishing to keep data confidential because of concerns about risk.

Where lives may hang in the balance, getting it right is crucial, says Budowle. The answers from phylogenetic forensics could mean sending an individual to prison or ostracizing a patient population. In cases involving bio-weapons, the conclusions could mean levying sanctions against a country or even going to war. Validating the tools and tests is all the more challenging in an area evolving nearly as fast as the microorganisms it traces. “It’s still an emerging field,” Budowle says. “We expect that what we are using today, we probably won’t be using two years from now.” ■

Shaoni Bhattacharya is a freelance science writer based in London.

1. González-Candelas, F., Bracho, M. A., Wróbel, B. & Moya, A. *BMC Biol.* **11**, 76 (2013).
2. Hanczaruk, M. *et al. Emerg. Infect. Dis.* **20**, 322–323 (2014).
3. de Oliveira, T. *et al. Nature* **444**, 836–837 (2006).
4. Scaduto, D. I. *et al. Proc. Natl Acad. Sci. USA* **107**, 21242–21247 (2010).