

# A Study of Human History Through the Lives of Disease-Causing Bacteria

Phylogenetics compares the genomes of extant organisms to reconstruct their ancestries, and their impact on humans. This is not very different from comparing languages and finding common or different origins between them.

## Aswin Sai Narain Seshasayee

We live in an era of big data in everything, including biology. The most prominent arm of big science in biology is genomics, the science of reading the entire genetic material of an organism, most touted for its ability to predict predisposition to disease, identify targets for personalised medicine, and rather tentatively even diet. And, of course, for its potential to add to the arsenal that we have to probe ancient history. Yes, ancient history.

Since the first full genome – that of a bacteriophage – was known in the 1970s, genomics has taken great strides culminating in the publication of the human genome, a project that took fifteen years and the toils of hundreds of scientists. This project was accompanied by the sequencing of the genomes of all sorts of other organisms, including bacteria, fungi, flies and mice.

What became apparent over the course of this journey was the fact that a single genome sequence, on its own and in the absence of other knowledge of biology and genetic sequences, is not terribly useful. But add more genetic data from other organisms to the mix and the genome sequence becomes a magic wand. One area that best illustrates the power of analysing and interpreting multiple genome sequences together is what is known as phylogenetics – comparing genomes, using the principles of evolution, to trace ancestries and origins.

A fundamental tenet of evolution is that the genetic pool of life on Earth changes over time, often in response to their changing circumstances. This is reflected in genome sequences. An oft quoted example is the rapid mutability of the flu virus: it changes so rapidly that even immunisation against this virus is seasonal. Genome sequences of these viruses year to year will show this variation. Similarly, variation among humans and between, say, humans and chimps are also reflected in their genomes.

Phylogenetics often compares the DNA or genome sequences of extant organisms to describe relationships between them, and also attempt to reconstruct their ancestries. This is not very different from comparing languages and finding common or different origins among them. The one crucial difference is the better

tractability of DNA sequences, whose lexicon is a four-letter alphabet across all known forms of life. Our ability to reconstruct ancestries gets better when the organisms being compared are more closely related to each other. As the evolutionary distance between two compared genomes increases, the molecular events that resulted in their divergence become murkier and the confidence in building accurate relationships between them. Knowing what their ancestors might have looked like genetically becomes harder.

In recent times, new techniques and technologies have permitted not only the rapid and economic sequencing of the genomes of a large number of organisms, but also allowed us to dig (literally) the genetic material of long-lost creatures from fossils and other remains. In other words, we now have the means to directly sequence the DNA of some of our ancestors long lost. For example, a few years ago, scientists sequenced the genome of the 38,000 year old remains of a Neanderthal woman. Comparison of the neanderthal genome with those of humans helped find evidence for interbreeding between humans and the Neanderthals before human migration into Eurasia. Thus, the Neanderthals appear to have contributed to the genetic material of modern humans. This is one prominent example of DNA sequences being used to study the distant past.

History is a murky subject, and often dominated by the influences of prominent humans on the course of events. Humans never lived alone. They have lived not only among animals and plants but also among a vast excess of microbial life, the unseen backstage players that – well beyond being also-rans – have often changed the course of history. Beside the fact that there would be no human life as we know it without microbial life, the plethora of infectious diseases that microbes cause have irreversibly altered the course of history. These have been elegantly traced in at least two books: William McNeill's *Plagues and Peoples* (Anchor Books) and Irwin Sherman's *Twelve Diseases that Changed our World* (ASM Press).

The most prominent example of the indelible impact of disease on human history is what might have been a smallpox epidemic that killed many Aztec warriors, who earlier had been successful in warding off an attack of the Spanish Conquistadors. The immunological naivety of the South American people eventually resulted in a rapid spread of the contagion killing over a quarter of the population. As highlighted by McNeill, “the psychological implication” of an epidemic that killed their own people leaving the Spanish invaders untouched is “worth considering”. This could only be explained by the “supernatural”, and the “superior power of the Gods that the Spaniards worshipped”. This in turn would have paved the way for the establishment of Christianity among the South Americans!

A second great epidemic, which ironically might have helped slow down the advance of the form of Christianity that produced the Dark Ages of Europe was the plague. The plague infection that caused the 14th century's Black Death ravaged

Europe and erased millions of people from the population registers. Many priests who performed the last rites on the victims of the plague were also consumed by the disease, presumably causing widespread scepticism over the power of their religious fervour in fighting off an 'evil' infection.

Now, it is thought that there have been at least three great plague epidemics (referred to as pandemics to indicate their global impact). The first hit the sixth century Eastern Roman empire during the rule of the emperor Justinian. The second was Black Death, and the third being a 19th-20th century pandemic probably of Chinese origin. Europe had its many bouts of localised plague epidemics till the 18th century. India had a taste of plague when it ran amok in Surat for two weeks in 1994, displacing a hundred thousand people.

Today, we know that the plague is a disease caused by the bacterium *Yersinia pestis*, which infects rodents and is transmitted among them and to humans by blood fleas. In fact, the discovery of the causative agent of the disease and the vehicle of its spread, under the watchful eyes of Louis Pasteur and Robert Koch, is a cornerstone in the history of infectious disease biology. Now, an engaging story of the genetic ancestry of the plague bacterium doing the rounds today, and its link to the first two great plague epidemics, is emerging thanks to the advent of genomics.

In 2011, Kirsten Bos – then at the McMaster University in Canada and now at the Max Planck Institute for the Science of Human History at Jena in Germany – and coworkers sequenced the genome of *Yersinia pestis* isolated from four exhumed victims of the Black Death. By comparing this historic genome sequence with the DNA sequence of currently circulating versions of the bacterium, these researchers found a staggering result. The causative agent of the Black Death and the extant versions of *Yersinia pestis* differed at less than a 100 letters, out of the 4.5 million letter long DNA sequence of this bacterium. This squarely placed the agent that caused the Black Death as the ancestor of the currently circulating variants of the bacterium. In fact, the genome of the Black Death bacterium differed from that of the inferred common ancestor of the extant *Yersinia pestis* at only two letters.

Later work from Bos, sequencing the plague bacterium derived from later European epidemics, showed that the ancestral *Yersinia pestis* that caused Black Death spread through Europe, travelling to Asia and finally nucleating the later pandemic in our continent. Finally, what of the plague of the Justinian Roman Empire? It was distinct from the agent of Black Death and is probably extinct or not sampled in extant populations, as found by another genomic study by Henrik Poinar, who was involved in Bos's first study of the Black Death, and his colleagues.

Getting to the today and now, rapid genome sequencing has enabled tracing the origins of raging epidemics and pandemics, and at times within timescales that have allowed for the implementation of effective containment procedures. As DNA

sequencing becomes cheaper and more accessible, the day may not be too far when a genome sequencer will be part of the toolkit of a standard diagnostic laboratory in the country.

*Aswin Sai Narain Seshasayee runs a laboratory researching bacterial biology at the National Centre for Biological Sciences, Bengaluru. Beyond science, his interests are in classical art music and history.*