GETTY

The bacterium behind the plague has now been traced back centuries.

GERMS, GENES AND SOIL: TALES OF PATHOGENS PAST

Archaeogeneticists are turning their attention towards ancient microbes to give bacteria their due in human history. **By Amber Dance**

The first documented epidemic of syphilis broke out in the French army during its invasion of Naples in 1495. From there, the sexually transmitted disease spread quickly across Europe — but where did it come from?

The timing of the outbreak, among other evidence, has led some researchers to hypothesize that the crews of the Italian explorer Christopher Columbus carried the bacterium behind the disease — *Treponema pallidum* — back from the Americas during transatlantic crossings (ref. 1). But fresh data from ancient microbial genomes raise doubts about that theory, says Kerttu Majander, an archaeogeneticist at the University of Basel in Switzerland.

When sampling bones and teeth from people buried in Europe around the time of and in the centuries after Columbus's voyages, Majander discovered two cases of syphilis, one instance of a related disease called yaws and one infection with a related bacterium that is no longer circulating[2]. These conditions were all due to subspecies of *T. pallidum*, and the genetic diversity seen in the infected skeletons — found in the Netherlands, Finland and Estonia, all far from where Columbus docked — suggests that these microorganisms might have been circulating in Europe well before the 1490s. Whatever else Columbus and his crews did, they might be innocent of bringing syphilis to the Old World.

Beyond historical curiosity, the lives of microbes that coexist with people are closely intertwined with the stories of humanity in ways that are important to unravel. "The ongoing pandemic showed us very nicely that we really do need to understand key things about how infectious diseases emerge and evolve over time," says Sebastian Duchene, a computational biologist working on the evolution of infectious diseases at the Doherty Institute in Melbourne, Australia.

The recovery of sequenceable shreds of DNA from centuries-old skeletons seemed impossible just a couple of decades ago, says Kirsten Bos, an ancient-DNA scientist at the Max Planck Institute for Evolutionary

Anthropology in Leipzig, Germany. But next-generation sequencing is perfectly suited to the short strands that can be recovered from ancient microbial genomes, some of which are just 30 base pairs long.

The techniques have allowed investigators to assign pathogens to diseases that were described only vaguely in historical records, and could be used to trace human activities and migrations on the basis of the microbes that accompanied them. Bos and other researchers have confirmed that the plague bacterium, *Yersinia pestis*, was indeed behind the second plague pandemic from the fourteenth to eighteenth centuries[3], the plague of Justinian that began in the sixth century[4]; and even epidemics that pre-date any written plague records. The ancient genomes even suggest that the bacterium was infecting people before it acquired the ability to infect fleas, its current mode of transmission to humans[5]. Researchers have sequenced DNA from oral microbes going back tens of thousands of years[6], and might be able to push further still with the right samples.

The work requires careful sample preparation and powerful computational tools, plus the know-how to distinguish true ancient genomes from modern contaminants. There are also ethical questions that arise when scientists – mostly from wealthy Northern Hemisphere nations – publicize intimate biological details about ancient peoples in other parts of the world. Meanwhile, researchers continue to push technical boundaries, delving deeper into the intertwined history of humans and microbes.

"All these pathogens, they have their own stories," says Pooja Swali, a geneticist at the Francis Crick Institute in London. "We're all just trying to piece it together."

Seeking shredded DNA

Piecing together the stories of ancient microbes often starts with skeletal remains. A handful of diseases, including syphilis and leprosy, can leave their mark on bone, so scientists interested in those pathogens can extract DNA from a pea-sized sample. Otherwise, the inner cavity of teeth is a good place to look: the bloodstream feeds directly into dental pulp, which is generally well-protected over the centuries. Scientists interested in the oral microbiome might start with a chip of dental calculus, and preserved faeces, although rare, can provide clues to the gut microbiome. Even old food containers might contain microbes that provide clues about diet, says Christina Warinner, a biomolecular archaeologist at Harvard University in Cambridge, Massachusetts.

Whatever their source, ancient samples have been through the wringer compared with their counterparts in modern microbial genomics. Bones and teeth have often been sitting in soils, crypts or museums for generations. Water and heat damage DNA, so the colder and drier the conditions, the better. Even so, environmental microbes can soak into the bones over time. Modern microbes might also have become attached when remains were handled by archaeologists or stored in museums.

To avoid any further contamination when the samples are being prepared, ancient-DNA researchers work in heavy-duty clean rooms and use the cleanest possible reagents. Their labs are much like those used to study highly contagious pathogens, such as yellow fever and tuberculosis, but it's the samples, rather than the scientists, that are being protected. "We joke that we clean more than we do science," says Laura Weyrich, a palaeomicrobiologist at Pennsylvania State University in University Park.

Making matches

Sequencing the entire contents of a sample will yield a metagenome: snippets of genetic code from a multitude of organisms, including the human host, the microorganisms of interest and microbial contaminants. Maybe 1% will be the desired microbial DNA, even less if scientists are hunting for a particular pathogen, says Maria Spyrou, an archaeogenomicist at the University of Tübingen in Germany. But once researchers identify a specimen infected with a given bacterium, they can use small segments of the organism's genome as bait to fish out more of its genetic material.

"All these pathogens, they have their own stories. We're all just trying to piece it together."

Palaeo-microbial geneticists deal in rarities: sometimes there's only enough material present to sequence the genome once, says Meriam Guellil, a palaeogenomicist at the University of Vienna. That leaves a lot of room for error, and is why modern whole-genome sequencing relies on coverage of 30 times or more. With such patchy coverage for ancient genomes, "we have to be very careful about how we interpret our results", says Duchene.

Now that they have the sequences in hand, researchers can try to match them with those of known microbes. One popular approach involves a piece software called MALT[7], but owing to the intensive computing time involved, many members of the community are adopting faster algorithms, such as Kraken[8]. This program slices up genomes and samples into small segments of length *k* (called *k*-mers, about 30 bases long), and looks for matches between the two.

Both MALT and Kraken assume that the species in the ancient sample are represented in modern genome libraries, but that's a big assumption: only about 2% of the genomes of bacteria and other prokaryotes have been sequenced, according to one estimate[9].

An alternative is to skip the matching step and assemble the fragments into a genome from scratch. This requires deep sequencing coverage, but, according to Warinner, "it allows us to find species that haven't been described, and it allows us to find genes that haven't been described". For example, earlier this year, Warinner and her collaborators reported that they had reconstructed the genomes of 459 bacteria from the dental calculus of Neanderthals and *Homo sapiens* from as far back as 100,000 years[6].

Such analysis requires care and expertise, because red herrings abound. For example, it would be easy to confuse *Mycobacterium tuberculosis*, the bacterium behind tuberculosis, with one of its many close (and harmless) relatives that live in the soil that bones are often found in. One key clue to veracity is that ancient genomes should be badly damaged, whereas those of modern contaminants usually stand out as oddly pristine. That's because DNA slowly shatters over centuries, with single strands hanging loose at either end of the fragment. Often, amine groups can be lost from naked cytosine bases, and the cytosine turns into uracil – which is not usually found in DNA. In sequencing data, the uracils would show up as a thymines, the corresponding DNA base, at either end of fragments. A computational tool called mapDamage can help to track these patterns.

Ancient-genome specialists have built a collegial community that's eager to help those who lack this nuanced expertise. They call it SPAAM: Standards, Precautions, and Advances in Ancient Metagenomics. For example, says Swali, the SPAAM Slack workspace includes a channel called 'No Stupid Questions' for those looking to learn. Another SPAAM initiative, called AncientMetagenomeDir, catalogues published data on ancient genomes with all the relevant tags, such as database accession numbers and dates[10].

Pushing boundaries

Working with ancient DNA means collaborating not only with other molecular biologists, but also with museum curators, historians and even modern relatives of the people under study. Such collaborations can greatly enrich investigations, says Spyrou. For example, when investigating *Y. pestis* in two fourteenth-century cemeteries in Kyrgyzstan, Spyrou collaborated with Philip Slavin, a self-styled "scientist of the past" at the University of Stirling, UK. Slavin studied the tombstone epitaphs, diaries from nineteenth-century excavations and the community's trade connections with other parts of Eurasia, details that added important context to the genetics when the work was published in *Nature* in 2020 (ref. 11).

"It's not only about studying the genomes

of these pathogens or the genomes of these people," says Spyrou. "It's more about giving the complete picture."

That complete picture includes real people who once lived, worked, loved — and left descendants. Scientists sometimes forget that fact in their eagerness to publish their findings, warns María Ávila-Arcos, a palaeogenomicist at the National Autonomous University of Mexico in Juriquilla.

Ávila-Arcos is concerned that many labs undertaking the intensive, expensive work of ancient-DNA analysis are concentrated in wealthy nations. But some are studying samples from less-developed countries and reaping the rewards of publishing the historical narratives themselves[12]. "They are kind of extracting — I would use that word in some instances — ancient samples that are part of national patrimony and processing them in a factory setting to generate hundreds or thousands of ancient genomes, and getting these papers, without any real benefit to the local institutions or researchers or communities where the samples are taken from," she says.

Rather than swoop in to collect samples themselves, Ávila-Arcos says that scientists should collaborate with local researchers, who probably have a better understanding of the culture in the region. She also advises researchers to be careful about how they phrase their conclusions, to avoid stigmatizing certain groups. For example, her team studied the pathogens behind epidemics that occurred in the land now known as Mexico during the colonial era[13]. The genomes that she discovered were linked to strains of hepatitis B virus and parvovirus B19 that, the team found, were probably introduced to the country from Africa as a result of the transatlantic slave trade. In the paper, she was careful to emphasize that the link was with the trade, rather than implying that the epidemics should somehow be blamed on the individuals who were enslaved.

In studies involving more recent remains, relatives or descendants might have concerns about what findings should be published, says Weyrich. For instance, if the samples contain evidence of a sexually transmitted disease, the descendants might worry that publicizing that fact would stigmatize their family. It's important to confer with local people and their representatives before collecting samples and after obtaining results, she says. For a project involving analysing dental calculus in French Polynesia, for example, Weyrich has been working with organizations such as the Department of Culture and Heritage and the Museum of Tahiti and the Islands to identify shared goals and risks, to learn how locals would like to see the results published and to share early findings.

Another element of ethical practice is to consider publishing the raw data, says Warinner. That's important because it means scientists can get the maximum possible information from every bit of tooth or bone that is analysed in the lab, and no ancient material is wasted. Scientists interested in ancient human genomes might be tempted to discard data on microbial genes, for instance, but those could be of use to other researchers. Raw sequence data from a study of Bronze Age human populations[14], for example, was later used to identify seven cases of plague, extending the known relationship between humans and *Y. pestis* back to at least 5,000 years ago[5].

WERNER SIEMENS FOUNDATION, FELIX WEY

Dental calculus can preserve microbial DNA for millennia.

As scientists push the technical envelope, these ethical considerations will become more important. For example, researchers are now making headway with DNA-based viruses, such as hepatitis B (ref. 15) and smallpox (ref. 16), which have much smaller genomes than do bacteria. Viruses with fragile RNA genomes are even more challenging, but scientists have extracted RNA from preserved lung samples of both influenza[17] and measles[18] dating back more than a century.

"We're continually testing the boundaries of what we can actually access."

Researchers are even starting to use ancient sequences to resurrect ancient proteins. For instance, Warinner's team used the palaeolithic microbial genes they sequenced to create two proteins that generated previously unknown metabolites they called palaeofurans[6]. These molecules might have been involved in regulating bacterial photosynthesis, the authors speculate.

Bos says that every time she thinks ancient-DNA researchers have delved as deep into the past as molecular preservation will allow, a new paper proves her wrong. "I don't want to place any strong restrictions on how far back we can go," she now says. "We're continually testing the boundaries of what we can actually access in terms of ancient molecules that were preserved over time."

And every bit of information helps. For example, Majander is eager to learn more about the spread of syphilis and related pathogens during the early modern period. Physicians from the 1700s onwards characterized syphilis as a disease of 'sin-wracked' urban centres — was that true? The spread of the infection from mother to child during pregnancy could also reveal clues about family structures and the positions of women in the past, Majander thinks. Although only a handful of historical syphilis genomes are now available, more sequences could illuminate how the pathogen moved and influenced people's lives.

After all, microbes do not exist alone. Their history is often human history, and Majander and other scientists are uncovering it one genome at a time.

Amber Dance is a science journalist based in Los Angeles, California.

1. Rothschild, B. M. *Clin. Infect. Dis.* **40**, 1454–1463 (2005).
2. Majander, K. *et al. Curr. Biol.* **30**, 3788–3803 (2020).
3. Spyrou, M. A. *et al. Nature Commun.* **10**, 4470 (2019).
4. Wagner, D. M. *et al. Lancet Infect. Dis.* **14**, 319–326 (2014).
5. Rasmussen, S. *et al. Cell* **163**, 571–582 (2015).
6. Klapper, M. *et al. Science* **380**, 619–624 (2023).
7. Herbig, A. *et al.* Preprint at bioRxiv https://doi.org/10.1101/050559v1 (2016).
8. Wood, D. E. & Salzbert, S. L. *Genome Biol.* **15**, R46 (2014).
9. Zhang, Z., Wang, J., Wang, J., Wang, J. & Li, Y. *Microbiome* **8**, 134 (2020).
10. Fellows Yates, J. A. *et al. Sci. Data* **8**, 31 (2021).
11. Spyrou, M. A. *et al. Nature* **606**, 718–724 (2022).
12. Ávila Arcos, M. C., de la Fuente Castro, C., Nieves-Colón, M. A. & Raghavan, M. *Front. Genet.* **13**, 880170 (2022).
13. Guzmán-Solís, A. A. *et al. eLife* **10**, e68612 (2021).
14. Allentoft, M. E. *et al. Nature* **522**, 167–172 (2015). 15. Mühlemann, B. *et al. Nature* **557**, 418–423 (2018).
16. Duggan, A. T. *et al. Curr. Biol.* **26**, 3407–3412 (2016).
17. Patrono, L. V. *et al. Nature Commun.* **13**, 2314 (2022).
18. Düx, A. *et al. Science* **368**, 1367–1370 (2020).