



ANCIENT PROTEINS TELL THEIR TALES

*Where DNA degrades, proteins might persist.
So scientists are looking to fill in early human history
using some very old amino acids.*

BY MATTHEW WARREN

Some time in the past 160,000 years or so, the remains of an ancient human ended up in a cave high on the Tibetan Plateau in China. Perhaps the individual died there, or parts were taken there by its kin or an animal scavenger. In just a few years, the flesh disappeared and the bones started to deteriorate. Then millennia dripped by. Glaciers retreated and then returned and

retreated again, and all that was left behind was a bit of jawbone with some teeth. The bone gradually became coated in a mineral crust, and the DNA from this ancient ancestor was lost to time and weather. But some signal from the past persisted.

Deep in the hominin's teeth, proteins lingered, degraded but still identifiable. When scientists analysed them earlier this year,

they detected collagen, a structural support protein found in bone and other tissues. And in its chemical signature was a single amino-acid variant that isn't present in the collagen of modern humans or Neanderthals — instead, it flagged the jawbone as belonging to a member of the mysterious hominin group called Denisovans¹. The discovery of a Denisovan in China was a major landmark. It was the first individual found outside Denisova Cave in Siberia, where all other remains of its kind had previously been identified. And the site's location on the Tibetan Plateau — more than 3,000 metres above sea level — suggested that Denisovans had been able to live in very cold, low-oxygen environments.

But the finding also marked another milestone: it was the first time that an ancient hominin had been identified using only proteins.

It is one of the most striking discoveries yet for the fledgling field of palaeoproteomics, in which scientists analyse ancient proteins to answer questions about the history and evolution of humans and other animals. Proteins, which stick around in fossils for much longer than DNA does, could allow scientists to explore whole new eras of prehistory and use molecular tools to examine bones from a much broader part of the world than is currently possible, according to the field's proponents.

Previously, scientists had recovered proteins from 1.8-million-year-old animal teeth and a 3.8-million-year-old eggshell. Now, they hope that palaeoproteomics could be used to provide insights about other ancient hominin fossils that have lost all traces of DNA — from *Homo erectus*, which roamed parts of the world from about 1.9 million to 140,000 years ago, to *Homo floresiensis*, the diminutive 'hobbit' species that lived in Indonesia as recently as 60,000 years ago. By looking at variations in these proteins, scientists hope to answer long-standing questions about the evolution of ancient human groups, such as which lineages were direct ancestors of *Homo sapiens*. "I think that you can basically unlock the whole of the human tree," says Matthew Collins, a bioarchaeologist at the University of Copenhagen who has been at the forefront of the field since the 1980s, when it consisted of just a handful of researchers.

A COMING OF AGE

Despite the excitement, some argue that researchers could struggle to paint a definitive picture of human history from the information that researchers can get out of proteins, which is limited compared with that obtainable from DNA. And many worry that palaeoproteomics in general might be susceptible to spurious results, stemming from issues such as contamination. "You see very good research, and then you see people that publish things that are just very strange, because they don't think critically about the methods," says Philipp Stockhammer, an archaeologist at the Ludwig Maximilian University of Munich in Germany.

Over the past two decades, DNA retrieved

from ancient fossils has transformed scientists' understanding of human evolution. Analysis of the similarities and differences in the DNA of different hominin groups has allowed researchers to map out the tangled family tree in a way that was previously not possible. And genetic material has led to some major finds, such as the discovery of Denisovans in the first place.

But glaring gaps remain in that picture. DNA has been sequenced from just three groups of hominin: Neanderthals, Denisovans and *Homo sapiens*, mostly from specimens that are less than 100,000 years old (a notable exception is a pair of 430,000-year-old early Neanderthals from Spain²). Go a few hundred thousand years further back, and things get much murkier. This was a time period when a lot of exciting things were happening, says Frido Welker, a molecular anthropologist at the University of Copenhagen. It's when Denisovans and Neanderthals branched off from the lineage that would become modern humans, for example. But it remains a hazy part of human history. Researchers don't know, for instance, whether the ancient hominin *Homo heidelbergensis*, which lived around 700,000–200,000 years ago, was an ancestor of both *H. sapiens* and Neanderthals or part of only the Neanderthal branch, as some have suggested. "A lot of that happens beyond the reach of ancient DNA," says Welker.

Go back one million years or more, and things get even less clear. *H. erectus*, for example, first emerged in Africa around 1.9 million years ago, but without DNA evidence, it remains uncertain exactly how it is related to later hominins, including *H. sapiens*.

Ancient DNA has also left geographical blind spots. DNA degrades faster in warm environments, so although a 100,000-year-old specimen found in a cold Siberian cave might still harbour genetic material, a fossil that has spent that long in the heat of Africa or southeast Asia generally will not. As a result, little is known about the genetics of even relatively recent hominins from these regions, such as *H. floresiensis*.

Now researchers are hoping that protein analysis might begin to fill in some of those blanks. The idea is not new: as early as the 1950s, researchers had reported finding amino acids in fossils. But for a long time, the technology needed to sequence ancient proteins just didn't exist. "For most of my career, I honestly, genuinely believed that we would not be able to recover ancient protein sequences," says Collins.

That changed in the 2000s, after researchers realized that mass spectrometry — a technique used to study modern proteins — could also be applied to ancient proteins. Mass spectrometry essentially involves breaking down proteins into their constituent peptides (short chains of amino acids) and analysing their masses to deduce their chemical make-up.

Researchers have used this method to sift through hundreds of bone fragments to identify

the types of animal they came from. In this specific approach, called zooarchaeology by mass spectrometry or ZooMS, researchers analyse one kind of collagen. The mass of collagen's components differs in various groups and species, providing a characteristic fingerprint that allows researchers to identify the bone's source.

ZooMS was used in a 2016 paper³ to identify one hominin bone among thousands of fragments from Denisova Cave — a bone that DNA analysis would later show belonged to a hybrid individual, nicknamed Denny, with a Neanderthal mother and a Denisovan father. Even with that result alone, ancient protein analysis had already substantially expanded our view of human evolution, says population geneticist Pontus Skoglund at the Francis Crick Institute in London. Katerina Douka, an archaeologist at the Max Planck Institute for the Science of Human History in Jena, Germany, is now using the technique to search through 40,000 unidentified bone fragments from Asia in the hope of uncovering more ancient hominins.

"I THINK THAT YOU CAN BASICALLY UNLOCK THE WHOLE OF THE HUMAN TREE."

But ZooMS paints a picture only in broad brushstrokes. Once a bone is identified as belonging to a hominin, for example, other techniques are needed to delve deeper. So others have turned to shotgun proteomics, which aims to identify all the protein sequences in a sample — its proteome. The composition of the proteome depends on the kind of tissue being examined, but will often include various forms of collagen. This method spits out thousands of signals, which makes it much more informative than ZooMS, says Douka, but also trickier to interpret. By matching these signals to known sequences in databases, researchers can identify the exact sequences of collagen or other proteins in their sample.

Scientists can then compare this newly determined protein sequence to the same protein from other hominin groups, looking for similarities and differences in individual amino acids that will help to place the hominin on the family tree. This is similar to how ancient-DNA researchers look at single-letter variations in genetic sequences.

FILLING IN THE GAPS

Although researchers had used protein analysis alongside ancient DNA sequencing before⁴, the Tibetan Denisovan was the first ancient hominin for which proteins alone

H. FLORESIENSIS: P. BROWN ET AL. / NATURE; DENISOVAN TOOTH: R. REICH ET AL. / NATURE; DENNY HYBRID: TOM HIGHAM, UNIV. OXFORD; DENISOVAN JAWBONE: F. CHEN ET AL. (REF. 1) / NATURE; H. NALEDI: L. R. BERGER ET AL. / ELIFE; NEANDERTHAL: M. MEYER ET AL. (REF. 2) / NATURE; H. ERECTUS: NAT. HIST. MUS. ALANY.; STEPHANORHINUS: NAT. HIST. MUS. DENMARK; LUCY: 120 VIA WIKIMEDIA COMMONS; OSTRICH EGGSHELL: TERRY HARRISON

were analysed — and others could soon follow (see ‘Getting fossils to speak’). A look at the protein sequences from *H. heidelbergensis*, for example, could clarify its relationship to *H. sapiens* and Neanderthals.

Debates have swirled for a decade and a half over the nature of *H. floresiensis*, remains of which were discovered on the Indonesian island of Flores in 2003. Its relationship to other hominins is unclear, with suggestions that it could be a dwarf descendant of *H. erectus*, or perhaps even that it evolved from the *Australopithecus* genus that is more distantly related to modern humans. This group lived more than 2 million years ago, and counts the famous Lucy skeleton among its members.

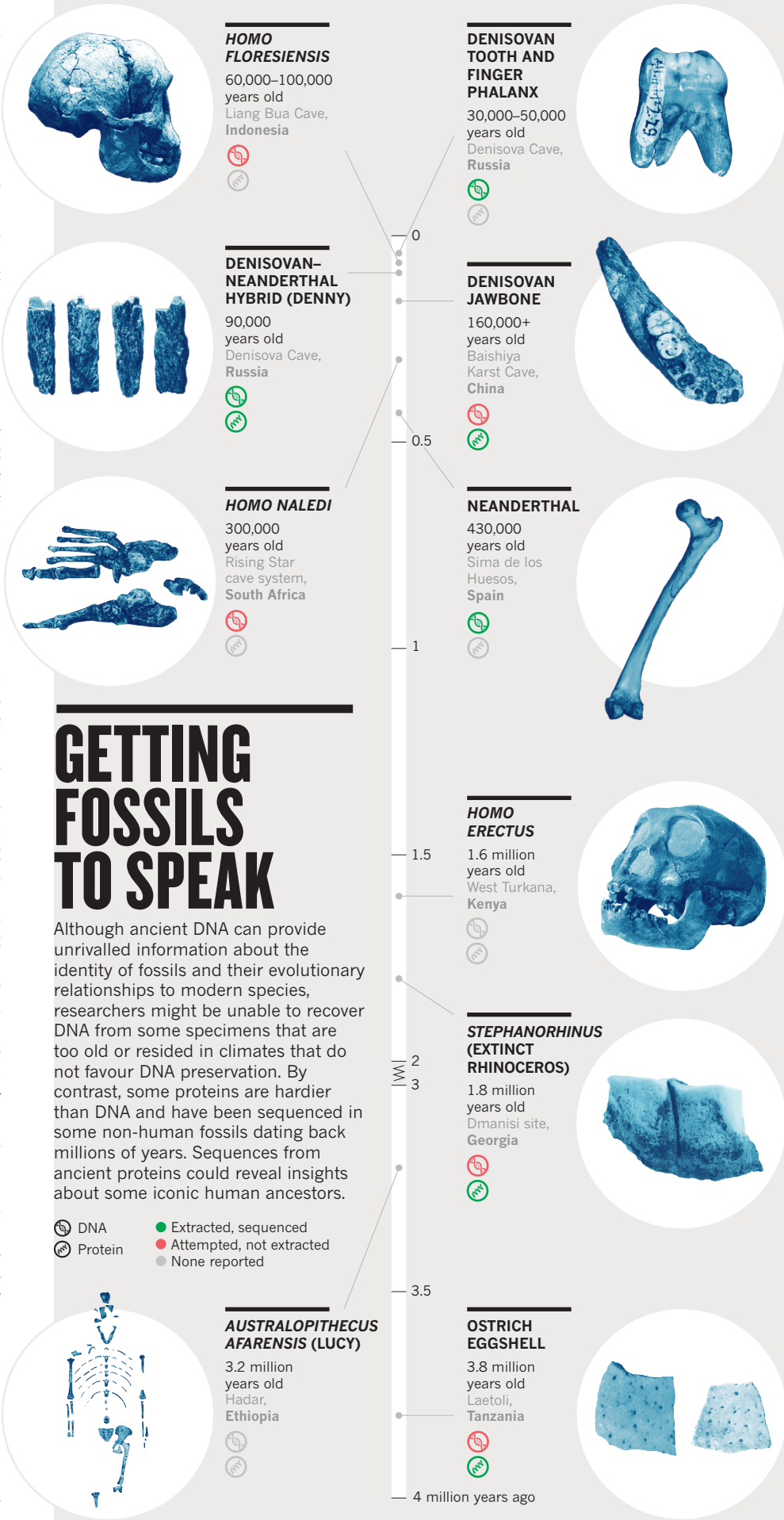
Proteomics could put that mystery to bed, says Collins. “I am utterly convinced that we have *Homo floresiensis* protein around, and it will be sequenceable, and it will tell us where that fits in the family tree,” he says. The same could be true of another small hominin, *Homo luzonensis*. Its bones and teeth were discovered in a cave on the island of Luzon in the Philippines several years ago, and reported on earlier this year⁵. Similarly to *H. floresiensis*, these samples have yielded no DNA. Armand Salvador Mijares, an archaeologist at the University of the Philippines in Quezon City, says that he is planning to send Welker an animal tooth from the cave where *H. luzonensis* was found, to test the viability of analysing proteins in ancient tropical materials.

As researchers prepare to do more proteomic analysis on ancient hominins, work on other animals is already revealing much about their evolutionary relationships in the deep past.

In a recent analysis, for example, Enrico Cappellini, a palaeoproteomics specialist at the University of Copenhagen, and his colleagues used proteomics to work out where the extinct rhinoceros *Stephanorhinus* fits on the rhino family tree. As reported in a preprint that has not yet been peer reviewed⁶, the team was able to extract proteins in remains from Dmanisi, Georgia, that were nearly 1.8 million years old. The pattern of amino-acid substitutions suggests that the animal was closely related to the extinct woolly rhinoceros (*Coelodonta antiquitatis*).

Whereas the proteins of the Tibetan Denisovan came from dentine, the bony tissue inside teeth, these *Stephanorhinus* proteins were locked away in the enamel that covers the tooth. This could be particularly useful for finding very old proteins, suggests Cappellini. Enamel is the hardest material in the vertebrate body and acts as what Cappellini calls a closed system, preventing amino acids from leaching out. The 1.8-million-year-old date “doesn’t represent a limit”, he says.

In fact, others have gone further back. Researchers have reported extracting collagen sequences from a



GETTING FOSSILS TO SPEAK

Although ancient DNA can provide unrivalled information about the identity of fossils and their evolutionary relationships to modern species, researchers might be unable to recover DNA from some specimens that are too old or resided in climates that do not favour DNA preservation. By contrast, some proteins are hardier than DNA and have been sequenced in some non-human fossils dating back millions of years. Sequences from ancient proteins could reveal insights about some iconic human ancestors.

DNA (blue circle with 'D')
Protein (green circle with 'P')

● Extracted, sequenced
 ● Attempted, not extracted
 ● None reported

3.4-million-year-old camel found in the Arctic⁷. And in a 2016 paper, Beatrice Demarchi, a biomolecular archaeologist at the University of Turin, Italy, and her colleagues extracted and sequenced proteins from a 3.8-million-year-old ostrich eggshell⁸. This shell wasn't preserved in a cold polar region: it came from a site in Tanzania, where the average annual air temperature is around 18 °C, says Demarchi. "You would not expect stuff to survive in such a hot environment," she says. Hominin proteins might be recoverable from the same places, she adds: "We've got to try, don't we?"

TEETHING PAINS

There are still hurdles to overcome before ancient proteins can bring the branches of the human evolutionary tree into focus. So far, researchers have been able to deduce the sequences of ancient hominin proteins fairly easily, because they already have DNA from Neanderthals, Denisovans and *H. sapiens*. This allows them to predict the protein sequences that are likely to appear in their mass-spectrometry signals. "You can identify fragments you expect to be there from known genome sequences, from either ancient organisms or present-day people, and look for them," says Svante Pääbo, a palaeogeneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

But as scientists look further back in time, they will need to work out the sequence of those amino acids without a map. That's an ongoing challenge for ancient proteomics, because proteins are degraded into small fragments, and samples are often contaminated with modern proteins, Pääbo says.

Collins is confident that it can be done. He points to a 2015 paper⁹ in which he, Welker and others mapped out the phylogenetic tree for South America's native ungulates, a varied group of peculiar-looking mammals that went extinct around 12,000 years ago. With no DNA available from ungulate fossils, the team had to sequence collagen proteins from scratch to compare them with those of other animals. They found that two extinct native ungulates, *Toxodon* and *Macrauchenia*, were closely related to a group that includes horses and rhinos — and not, as some researchers had thought, the group Afrotheria, which includes elephants and manatees.

Other limitations are more fundamental. Ancient teeth and bones contain a small number of proteins, so there are relatively few chunks of information that can be used to identify a specimen. Analysis of the Tibetan Denisovan, for example, revealed sequences from eight different kinds of collagen protein, totalling slightly more than 2,000 amino acids. Just one of these amino acids differed from Neanderthal and modern human sequences, identifying the sample as Denisovan. That means that even if a researcher were able to sequence the proteins

from a *H. erectus* specimen, for instance, there simply might not be enough information in the amino-acid sequences to say anything definitive about its relationship to modern or archaic humans. By comparison, a single ancient genome contains in the order of three million variants compared with any other genome, says Skoglund, and so is much more informative regarding evolution.

And because proteins often perform crucial functions — forming the structure of bone, say — they don't always change much as species evolve. Proteins that are specific to enamel, for instance, are exactly the same in Denisovans, *H. sapiens* and Neanderthals, so can't be used to distinguish between these groups. Welker says, however, that these proteins do vary in other great apes, and

“THIS TECHNIQUE IS SO INTERESTING AND SO FASCINATING AND IS REALLY GETTING A LOT OF ATTENTION.”

could be more informative when it comes to older hominin groups.

Still, researchers know very little about how protein sequences vary in populations of ancient humans. Scientists have sequenced only a single Denisovan genome, for example, which means that to identify the Tibetan Denisovan, the team compared the protein sequences to just one other member of that group. It could be that other Denisovans had different variants. "Many geneticists are quite sceptical of the methodology, but I think it's because they have come a long way in understanding genomic variation in ancient populations," says Douka.

LEARNING FROM THE PAST

There are other challenges, too. Some researchers are concerned that the broader buzz around palaeoproteomics could result in the field falling into the same traps as the ancient-DNA field did 20 years ago. Many apparently exciting results from the 1990s and early 2000s — the discovery of DNA from dinosaurs or insects trapped in amber, for example — later turned out to be false because they were products of contamination or other methodological errors. "I wouldn't be surprised if this happens to the proteomics world," says Douka.

Those leading the way in the field are aware

of these problems, and many researchers are making concerted efforts to create a robust science. Among them is Jessica Hendy, an archaeologist at the University of York, UK, who is pioneering the use of proteins to study the diet of early humans. In a 2018 paper, Hendy and her colleagues identified proteins in 8,000-year-old ceramics from Çatalhöyük in modern-day Turkey, which revealed that the ancient inhabitants ate various plants and animals, and even processed milk into whey¹⁰.

"This technique is so interesting and so fascinating and is really getting a lot of attention, especially right now," Hendy says. "We really need to be moving carefully," she adds. Together with Welker, Hendy is lead author on a paper outlining best practices for the field, from avoiding contamination to sharing data in public repositories¹¹.

Hendy adds that there needs to be more basic research into how proteins survive and degrade over long timescales. This kind of research might not make headlines, she says, but can give researchers much more confidence in their results. She points to Demarchi's work as an example: Demarchi found that the proteins in her 3.8-million-year-old eggshell had bound to the surface of the mineral crystals in the shell, essentially freezing them in place. "What's cool about that is that it's actually explaining why the proteins are surviving, which makes the finding so much more robust," says Hendy.

Even though there are still issues to sort out, progress in the field shows no signs of slowing. And whereas human evolution might get the most attention, scientists are using ancient proteomics in all kinds of ways, from studying markers of disease in the tartar of ancient teeth¹², to investigating which animal skins were used to create medieval parchments¹³.

Demarchi says she is excited by it all. And when it comes to working out the family trees of long-extinct organisms, she says, proteomics has the potential to make waves. "I don't think I'll see the end of it in my lifetime," she says. "It's going to be really quite big" ■

Matthew Warren writes for Nature from London, UK.

1. Chen, F. *et al.* *Nature* **569**, 409–412 (2019).
2. Meyer, M. *et al.* *Nature* **531**, 504–507 (2016).
3. Brown, S. *et al.* *Sci. Rep.* **6**, 23559 (2016).
4. Welker, F. *et al.* *Proc. Natl Acad. Sci. USA* **113**, 11162–11167 (2016).
5. Déroit, F. *et al.* *Nature* **568**, 181–186 (2019).
6. Cappellini, E. *et al.* Preprint at bioRxiv <https://www.biorxiv.org/content/10.1101/407692v1> (2018).
7. Rybczynski, N. *et al.* *Nature Commun.* **4**, 1550 (2013).
8. Demarchi, B. *et al.* *eLife* **5**, e17092 (2016).
9. Welker, F. *et al.* *Nature* **522**, 81–84 (2015).
10. Hendy, J. *et al.* *Nature Commun.* **9**, 4064 (2018).
11. Hendy, J. *et al.* *Nature Ecol. Evol.* **2**, 791–799 (2018).
12. Jersie-Christensen, R. R. *et al.* *Nature Commun.* **9**, 4744 (2018).
13. Fiddyment, S. *et al.* *Proc. Natl Acad. Sci. USA* **112**, 15066–15071 (2015).

CLARIFICATION

The News Feature 'Ancient proteins tell their tales' (*Nature* **570**, 433–436; 2019) did not make clear that the *Stephanorhinus* work cited in reference 6 was led by Enrico Cappellini.