

# Artificial Intelligence Study of Human Genome Finds Unknown Human Ancestor

The genetic footprint of a “ghost population” may match that of a Neanderthal and Denisovan hybrid fossil found in Siberia

By Brian Handwerk, smithsonian.com

February 8, 2019 8:30AM



Can the minds of machines teach us something new about what it means to be human? When it comes to the intricate story of our species' complex origins and evolution, it appears that they can.

A recent study used machine learning technology to analyze eight leading models of human origins and evolution, and the program identified evidence in the human genome of a “ghost population” of human ancestors. The analysis suggests that a previously unknown and long-extinct group of

hominins interbred with *Homo sapiens* in Asia and Oceania somewhere along the long, winding road of human evolutionary history, leaving behind only fragmented traces in modern human DNA.

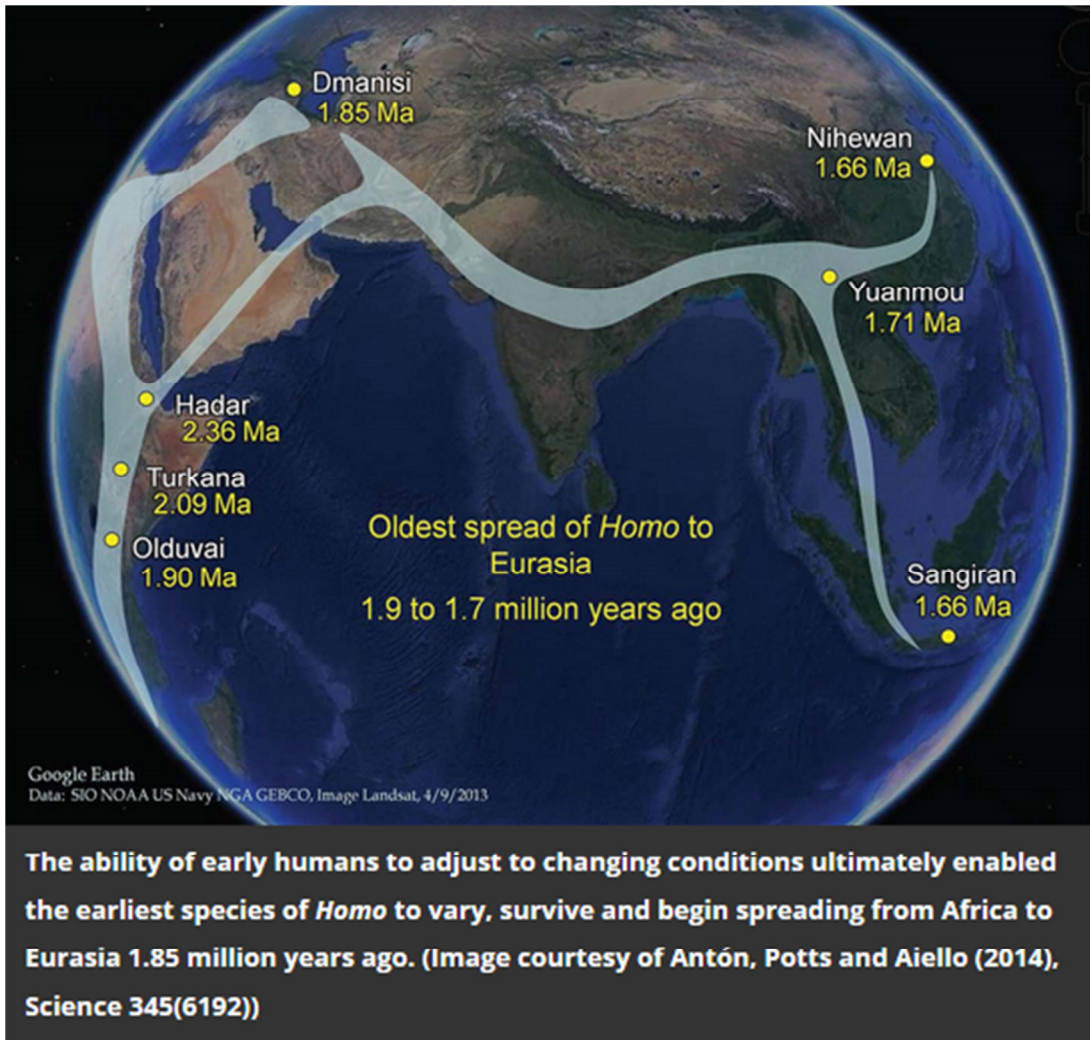
The study, published in *Nature Communications*, is one of the first examples of how machine learning can help reveal clues to our own origins. By poring through vast amounts of genomic data left behind in fossilized bones and comparing it with DNA in modern humans, scientists can begin to fill in some of the gaps of our species' evolutionary history.

<https://www.smithsonianmag.com/science-nature/artificial-intelligence-study-human-genome-finds-unknown-human-ancestor-species-180971436/#ppY5jy9gT2zYRh6U.99>

In this case, the results seem to match paleoanthropology theories that were developed from studying human ancestor fossils found in the ground. The new data suggest that the mysterious hominin was likely descended from an admixture of Neanderthals and Denisovans (who were only identified as a unique species on the human family tree in 2010). Such a species in our evolutionary past would look a lot like the fossil of a 90,000-year-old teenage girl from Siberia's Denisova cave.

Her remains were described last summer as the only known example of a first-generation hybrid between the two species, with a Neanderthal mother and a Denisovan father.

"It's exactly the kind of individual we expect to find at the origin of this population, however this should not be just a single individual but a whole population," says study co-author Jaume Bertranpetit, an evolutionary biologist at Barcelona's Pompeu Fabra University.



Previous human genome studies have revealed that after modern humans left Africa, perhaps 180,000 years ago, they subsequently interbred with species like Neanderthals and Denisovans, who coexisted with early modern humans before going extinct. But redrawing our family tree to include these divergent branches has been

difficult. Evidence for "ghost" species can be sparse, and many competing theories exist to explain when, where, and how often *Homo sapiens* might have interbred with other species.

Traces of these ancient interspecies liaisons, called introgressions, can be identified as places of divergence in the human genome. Scientists

observe more separation between two chromosomes than you'd expect if both of the chromosomes came from the same human species. When scientists sequenced the Neanderthal genome in 2010, they realized that some of these divergences represented fractions of our genome that came from Neanderthals. Studies have also revealed that some living humans can trace as much as 5 percent of their ancestry to Denisovans.

“So, we thought we'd try to find these places of high divergence in the genome, see which are Neanderthal and which are Denisovan, and then see whether these explain the whole picture,” Bertranpetit says. “As it happens, if you subtract the Neanderthal and Denisovan parts, there is still something in the genome that is highly divergent.”

Identifying and analyzing the many divergent places throughout the genome, and computing the countless genetic combinations that could have produced them, is too big a job for humans to tackle on their own—but it's a task that may be tailor made for deep learning algorithms.

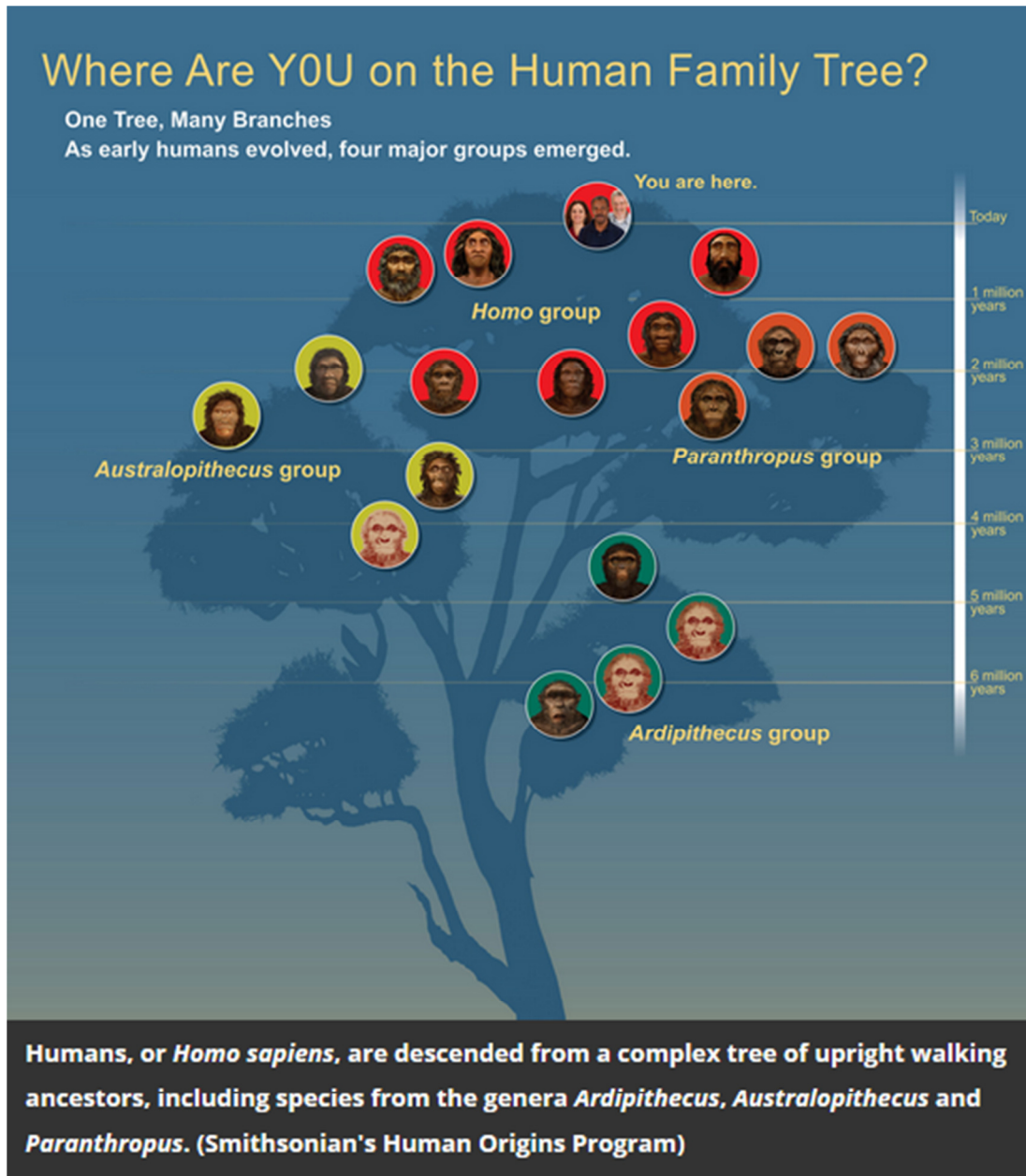
Deep learning is a type of artificial intelligence in which algorithms are designed to work as an artificial neural network, or a program that can process information the same way a mammalian brain would. These machine learning systems can detect patterns and account for previous information to “learn,” allowing them to perform new tasks or look for new information after analyzing enormous amounts of data. (A common example is Google DeepMind's AlphaZero, which can teach itself to master board games.)

“Deep learning is fitting a more complicated shaped thing to a set of points in a bigger space,” says Joshua Schraiber, an evolutionary genomics expert at Temple University. “Instead of fitting a line between Y and X, you're fitting some squiggly thing to a set of points in much bigger, thousand-dimensional space. Deep learning says, ‘I don't know what squiggly shape should fit to these points, but let's see what happens.’”

In this case, machines were set to work analyzing the human genome and predicting human demographics by simulating how our DNA might have evolved over many thousands of possible scenarios of ancient evolution. The program accounted for the structure and evolution of DNA as well as models of human migration and interbreeding to try to fit some of the pieces together in an incredibly complex puzzle.

The researchers trained the computer to analyze eight different models of the most plausible theories of early human evolution across Eurasia. The models came from previous studies that attempted to come up with a scenario that would result in the current picture of the human genome, including its known Neanderthal and Denisovan components.

“There could be other models, of course, but these models are the ones that other people have been proposing in the scientific literature,” Bertranpetit says. Each model begins with the accepted out-of-Africa event, then features a different set of the most likely splits between human lineages, including various interbreedings with both known species and possible “ghost” species.



“With each of these eight models, we calculate over weeks of computations how well they are able to reach the actual, present genetic composition of humans,” Bertranpetit says. “Every time we do a simulation, it’s a simulation of a possible path of human evolution, and we have run those simulations thousands of times, and the deep learning algorithms are able to recognize which of the models best suit the data.”

The machine’s conclusion? An ancestor species is present in our lineage that we have yet to identify. “By far, the only models we tested that really are

backed by the data are the ones having this ghost population introgression,” Bertranpetit says.

The intriguing study and others like it may help redraw the map of how humans migrated and evolved through what appears to be an increasingly complicated ancient world in Eurasia and Oceania.

“It’s certainly interesting and consistent with the emerging picture of a complex reticulated phylogeny in human evolution,” Iain Mathieson, a University of Pennsylvania population geneticist, says via email. “I’m not even sure it makes sense to

talk about 'introgression events' when that seems to be the norm." In fact, because only eight models were tested and many others could be possible, Mathieson adds that the new findings are "certainly a plausible scenario, but the reality is likely even more complex."

As new fossil discoveries are made in the field, updated models can now be tested against the human genome using these types of programs. Schraiber says the power of deep learning for studying human origins lies precisely in its capability to analyze complex models.

"If you want to do an extremely detailed model because you're an anthropologist, and you want to know if this introgression happened 80,000 years ago or 40,000 years ago, that's the power of a deep learning approach like this."

Complex as they are, the interbreedings of ancient Eurasia are still only one part of our human story. Bertranpetit believes that future advances in deep learning can help uncover other new chapters.

"This kind of method of analysis is going to have all kinds of new results," he says. "I am sure that people working in Africa will find extinct groups that are not recognized yet. No doubt Africa is going to show us surprising things in the future."

"With each of these eight models, we calculate over weeks of computations how well they are able to reach the actual, present genetic composition of humans," Bertranpetit says. "Every time we do a simulation, it's a simulation of a possible path of human evolution, and we have run those simulations thousands of times, and the deep learning algorithms are able to recognize which of the models best suit the data."

The machine's conclusion? An ancestor species is present in our lineage that we have yet to identify. "By far, the only models we tested that really are

backed by the data are the ones having this ghost population introgression," Bertranpetit says.

The intriguing study and others like it may help redraw the map of how humans migrated and evolved though what appears to be an increasingly complicated ancient world in Eurasia and Oceania.

"It's certainly interesting and consistent with the emerging picture of a complex reticulated phylogeny in human evolution," Iain Mathieson, a University of Pennsylvania population geneticist, says via email. "I'm not even sure it makes sense to talk about 'introgression events' when that seems to be the norm." In fact, because only eight models were tested and many others could be possible, Mathieson adds that the new findings are "certainly a plausible scenario, but the reality is likely even more complex."

As new fossil discoveries are made in the field, updated models can now be tested against the human genome using these types of programs. Schraiber says the power of deep learning for studying human origins lies precisely in its capability to analyze complex models.

"If you want to do an extremely detailed model because you're an anthropologist, and you want to know if this introgression happened 80,000 years ago or 40,000 years ago, that's the power of a deep learning approach like this."

Complex as they are, the interbreedings of ancient Eurasia are still only one part of our human story. Bertranpetit believes that future advances in deep learning can help uncover other new chapters.

"This kind of method of analysis is going to have all kinds of new results," he says. "I am sure that people working in Africa will find extinct groups that are not recognized yet. No doubt Africa is going to show us surprising things in the future."

