

Harvard Magazine  
Published on *Harvard Magazine* (<https://www.harvardmagazine.com>)

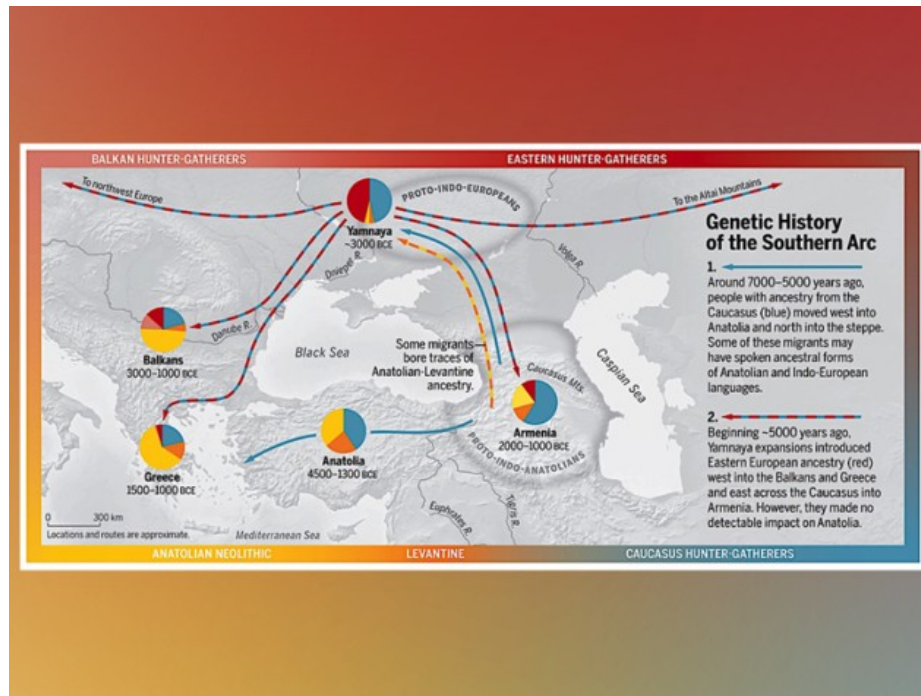
[Home](#) > Seeking the First Speakers of Indo-European Language

Research

## Seeking the First Speakers of Indo-European Language

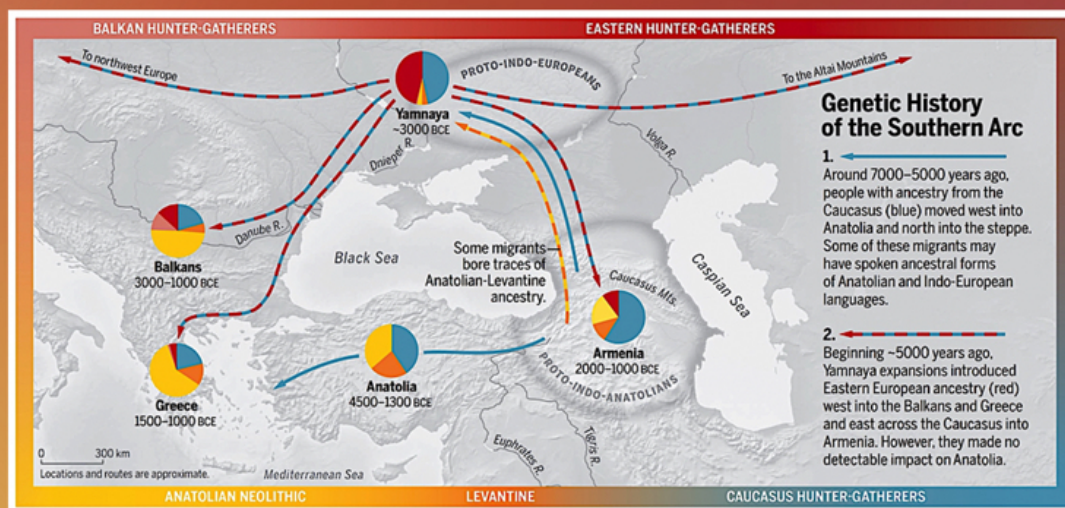
by [Jonathan Shaw](#) <sup>[1]</sup>

8.25.22



[More than 5,000 years ago, Caucasus hunter-gatherers from the highlands between the Black and Caspian Seas traveled west to Anatolia and north to the steppe, splitting their Proto-Indo-European language into two branches. From the steppe, their Yamnaya horse-herder descendants spread their language and genes into daughter languages and cultures across Eurasia. Border colors indicate the geographic origins of five source populations before their migrations \(shown by correspondingly colored arrows\), while the pie charts show the post-migration admixtures in these regions.](#)

[Figure reprinted with permission from I. Lazaridis et al., \*Science\* 377:939\(2022\).](#)



More than 5,000 years ago, Caucasus hunter-gatherers from the highlands between the Black and Caspian Seas traveled west to Anatolia and north to the steppe, splitting their Proto-Indo-European language into two branches. From the steppe, their Yamnaya horse-herder descendants spread their language and genes into daughter languages and cultures across Eurasia. Border colors indicate the geographic origins of five source populations before their migrations (shown by correspondingly colored arrows), while the pie charts show the post-migration admixtures in these regions.

Figure reprinted with permission from I. Lazaridis et al., *Science* 377:939(2022).



[2]

[3] [3] [3] [3]

A new study of ancient DNA from 727 individuals who lived in the regions cradling the southern half of the Black Sea, and extending into the Levant and western Iran, narrows the hunt for the origins of Indo-European languages—spoken today as a first language by almost half the world’s population. The research also documents genetic homogenization and stability among the population of farmers living between about 15,000 and 7,000 years ago in what is now Turkey, sheds new light on how an early form of Indo-European language may have spread in ancient Greece, and reveals the surprising discovery that the ancestry of the population of Rome during the Imperial period was drawn principally from Anatolia. These findings are the result of a 206-person collaboration led by staff scientist Iosif Lazaridis of the David Reich lab at Harvard, and by Songül Alpaslan-Roodenberg of the Reich lab and the Ron Pinhasi lab at the University of Vienna (Reich, a professor of genetics and of human evolutionary biology, and Pinhasi, an associate professor of evolutionary anthropology, are co-senior authors of the three related studies [published today in](#) <sup>[4]</sup>*Science* <sup>[4]</sup>.) The group’s work more than doubles the amount of ancient DNA from this region, and extends Reich’s pioneering studies of early human origins forward into periods for which there start to be scattered historical records.

Indo-European languages are the first language of more than 3 billion people in Europe, across northern India, the Iranian plateau, and as far east as Siberia (and on other continents as a result of colonialism, including in the United States). Beginning almost 500 years ago, scholars began to notice similarities between languages such as Sanskrit and Latin, and as the field of linguistics matured, it became clear that hundreds of such languages were connected by common root words. But where and when did the original language arise, and who spoke it?

Answering such questions has in the past been principally the work of archaeologists, linguists, and physical anthropologists. But more recently, as the analysis of ancient DNA has improved—aided by the 2015 discovery that DNA in the petrous bone of the inner ear can survive for millennia even in warm climates—geneticists, collaborating with experts in material culture and language, are making important contributions to the study of human history. (For a recent feature on the development of this science, see [“Telling Humanity’s Story through DNA”](#) <sup>[5]</sup>.)

Among the startling discoveries of the past decade has been that Indo-European languages seem *not* to have been spread by Anatolian farmers living in what is now Turkey, as was commonly thought, but rather by horse-herding nomads who lived on the Eurasian steppe, a people called the Yamnaya. A host of linguistic evidence suggesting this possibility was first marshaled persuasively by archaeologist David Anthony in his 2007 book, *The Horse, the Wheel and Language: How Bronze Age Riders from the Eurasian Steppe Shaped the*

*Modern World.* “I made the right guesses,” says Anthony modestly, now working with Reich in retirement as an associate of the department of human evolutionary biology.

In 2015, genetic evidence published by Reich and colleagues proved Anthony was on target. They showed that the Yamnaya spread more than language throughout Eurasia: beginning about 5,000 years ago, their genes began to appear everywhere from northern Europe to the Indian subcontinent.

But the Yamnaya are not thought to have invented the Indo-European language that they spoke—only to have spread it. Where could it have come from? Whole genome analysis, paired with insights from linguists, now points to an answer to this question, too.

## Support from Genetic Evidence

- *The ancient Indo-European languages spoken in Anatolia and on the steppe appear to have split from a common proto-language.*
- *Anatolia was genetically isolated after this split.*
- *The Anatolian and steppe speakers of these early Indo-European languages share a common ancestry somewhere in West Asia.*

In Anatolia, as if in isolation from its neighbors, the new research reveals almost no trace of steppe (Yamnaya) ancestry in samples of ancient DNA. Yet Hittite, a now extinct Indo-European language, was spoken there. Linguistic evidence (such as a lack of root words for wheeled vehicles) suggests that Hittite and the language spoken by the Yamnaya might have split early in the evolution of Indo-European languages from a common ancestral tongue.

“Ancient DNA data, building on decades of research in physical anthropology and archaeology,” says Reich, “is contributing to a qualitatively richer and more comprehensive picture of the origins of the first farmers.” In Anatolia, the first farmers descended from inhabitants of the Levant, the “fertile crescent” where agriculture first arose 11,000 years ago. Subsequent migrants to the region from West Asia mixed with this population, and continued to do so *within* Anatolia in what the researchers describe as a process of homogenization. “Anatolia was home to diverse populations descended from both local hunter-gatherers and eastern populations of the Caucasus, Mesopotamia, and the Levant,” says Alpaslan-Roodenberg. But the subsequent “homogenization in Anatolia” of these groups over time was coupled with “impermeability” to genes coming in from Europe or the steppe, the researchers found. The ancestry of Anatolian farmers shifted gradually over time as part of an insular, *intra-Anatolian* admixture up through the medieval period. By then, most farmers there were descended primarily from ancestors in the Caucasus, the region between the Black and Caspian seas including present-day Armenia, Azerbaijan, Georgia, and parts of southern Russia.

These genetic findings support the linguistic theory of an ancestral language common to Anatolians and the Yamnaya; and they explain how Anatolian languages persisted independently thereafter due in part to the relative genetic isolation of the region’s population from the rest of Europe. (Anatolians, the researchers also discovered to their surprise, contributed the majority of DNA to the peoples of the Roman Empire, as well as to the population of the city of Rome itself.)

The researchers do include a caveat: “in contrast to findings about movements of people,” they write, “the relevance of genetics to debates about language origins is more indirect because languages can be replaced with little or no genetic change, and populations can migrate and mix with little or no linguistic change. Nevertheless,” they continue, “the detection of migration is important because it identifies a plausible vector” for shifts in language.

## Origins of the Yamnaya

- *Analysis of the new genetic data reveals that the Yamnaya and Anatolian peoples share a common ancestry in the highlands of West Asia.*

If the Yamnaya did not invent the proto-Indo-European language that they spread from the steppe beginning about 5,000 years ago, where did it originate? Could ancestors of the Yamnaya have brought the language to the steppe in a migration, and if so, where did they come from, and when? The researchers discovered evidence of two such migrations in their data, in the form of two gene flows into the steppe from two different groups, both with origins in West Asia. Either one, the researchers write, “may have induced linguistic change there.” The researchers found that from 35 to as much as 50 percent of Yamnaya ancestry—what they characterize as a “substantial contribution”—came from the south, specifically the South Caucasus-Zagros area. Critically, the discovery links “the Proto-Indo-European-speaking Yamnaya with the speakers of Anatolian languages”; both share ancestry in the highlands of West Asia (the Middle East, including the Caucasus and Zagros mountains).

What is needed now, the geneticists write, is “a concrete research program of investigating the archaeological cultures of West Asia, the Caucasus, and the Eurasian steppe to identify a population driving transformations of both the steppe and Anatolia, linking the two regions.” None of the individuals sampled in the current study fit the genetic profile, which includes substantial hunter-gatherer ancestry. “The discovery of such a ‘missing link’ (corresponding to Proto-Indo-Anatolians if our reconstruction is correct),” they write, “would bring to an end the centuries-old quest for a common source binding through language and some ancestry many of the peoples of Asia and Europe.”

## Completing the Arc of Indo-European Expansion

- *Some men living in Armenia today are direct patrilineal descendants of the Yamnaya.*
- *In Greece, traces of the genes of steppe peoples suggest they integrated with the locals, rather than replacing them, raising new questions about how Indo-European languages spread there.*

The study also led to further insights into the Yamnaya expansion, because the researchers sequenced ancient DNA from regions that had previously not been well represented, especially in Armenia to the east of the Black Sea, and in the Balkans to the west. They found strong evidence of patrilineal descent from the Yamnaya people in Armenia, extending all the way to the present day: “men whose fathers’ fathers across thousands of years can be directly linked to the earliest Indo-Europeans,” says Reich.

In southeastern Europe, by contrast, the ancient inhabitants exhibit “extraordinary heterogeneity” in their ancestry; “a picture emerges” the researchers write, “of a fragmented genetic landscape that may well parallel the poorly understood linguistic diversity” of the region. It appears that the steppe-descended speakers who introduced early Indo-European language to the area did not run roughshod over the native inhabitants, as they did in what is now Germany, and in Britain, where 90 percent of the native population was replaced by steppe-descended peoples. In early Greece, for example, “steppe ancestry was common at low levels [of about 10 percent] in both elite [Mycenean] and non-elite [Minoan] individuals,” says Lazaridis. “Some elite men traced their paternal descent to steppe populations, but others, like the famous Griffin Warrior near ancient Pylos from whom we recovered DNA, did not have any steppe ancestry at all. We have to imagine steppe migrants as a population element that became integrated, both socially and genetically, into Aegean societies, and not as a people apart that dominated them.”

How then, did early Indo-European language become established in this region, if it wasn't through domination of the kind that characterized steppe expansion in northern Europe? Perhaps Indo-European functioned as a “lingua franca,” the researchers suggest, “facilitating communication among speakers of the diverse languages of previous farmer and hunter-gatherer populations” who populated the Balkan peninsula in the ancient world.

**Source URL:** <https://www.harvardmagazine.com/2022/08/indo-european-languages?page=all>

#### Links

- [1] <https://www.harvardmagazine.com/profile/Jonathan-Shaw>
- [2] <https://www.harvardmagazine.com/print/81689?page=all>
- [3] <https://www.addthis.com/bookmark.php?v=300>
- [4] <https://www.science.org/doi/10.1126/science.abm4247>
- [5] <https://www.harvardmagazine.com/2022/07/feature-ancient-dna>