

Beating the retreat from the Steppe hypothesis

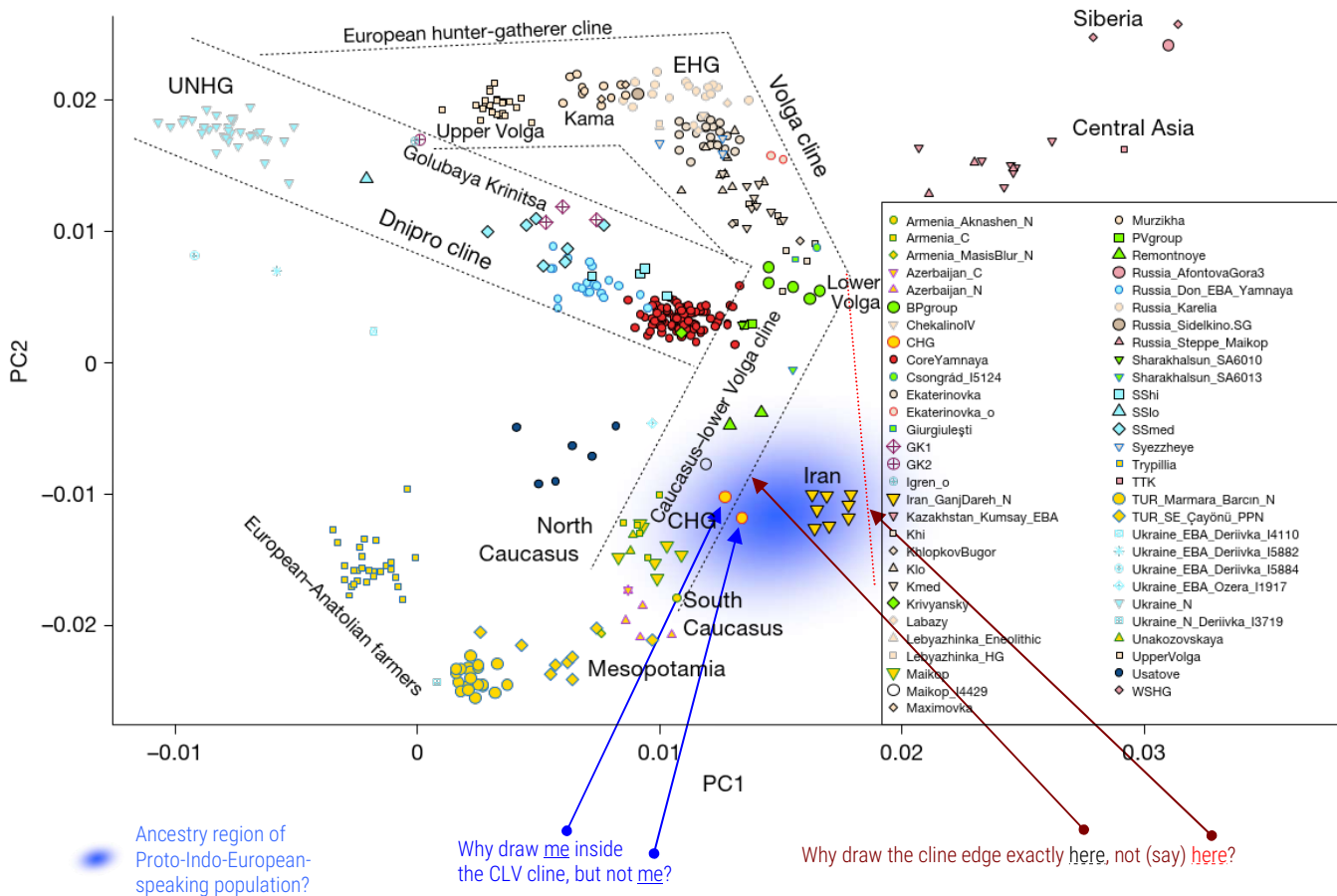
Commentary by Paul Heggarty on:

Lazaridis *et al.* (2025): 'The genetic origin of the Indo-Europeans' <https://doi.org/10.1038/s41586-024-08531-5>

Nature has just released (5th February 2025) a paper on "The genetic origin of the Indo-Europeans", by Lazaridis *et al.* (2025). Indo-European is a linguistic concept, the name of a family of languages, but this paper presents no language data or analyses. Rather, it reports data and analyses of ancient DNA.

Although presented as if supporting the Steppe hypothesis of Indo-European origins, the paper's basic result is actually a retreat from it. Some smoke and mirrors are deployed to cover this retreat, not least in proposing to change what 'Indo-European' actually refers to (page 8). Harvard's shot was off-target, so they propose moving the goalposts. (See below on this naming issue.) This still does not obscure where the whole language family originated: not on the Steppe, as this paper itself reconfirms, only not in such clear terms. Instead, it focuses on asserting the Steppe as the home of *most* of the family, through some self-citation, but introduces no new data on that, and no linguistic data at all.

Three years ago, Lazaridis *et al.* (2022) themselves acknowledged that the family's ultimate origins lie where many have long argued: in "the highlands of West Asia, the ancestral region". On page 1 of this new paper, they further subscribe to it being "widely agreed" that a major ancestry component is "Neolithic people from [the] Zagros and south Caucasus". Note that Neolithic here entails farmers (crops and animals), and that the Zagros mountains form the "hilly flanks" (Braidwood 1960) to the Fertile Crescent, not least in north-western Iran. The corresponding aDNA samples in the new paper's Figure 1b (reproduced below, with some observations) are labelled 'Iran', and in the key specifically 'Iran Ganj Dareh N[eolithic]'. The link with (some) early farming origins is clear.



Some observations on Figure 1b in Lazaridis *et al.* (2025)

As always, it is good that that this paper brings further aDNA coverage of this key region (just as Ghalichi *et al.* 2024 did in *Nature* a few months ago, though curiously not cited here). The new coverage brings refinements on the key role of this particular ancestry component, taken slightly more broadly in much other recent work as 'CHG/Iranian'. The new data reconfirm that it is essentially this component, from this region, that expands. One movement heads northwards to become the main ancestry component of core Yamna. The paper blurs this direction over time by its presentation of a 'cline' that it calls "Caucasus–lower Volga" (CLV), but the key population movement is spreading from the Caucasus end and heading northwards, not from the lower Volga southwards. The CLV cline itself is questionable: it has very few samples in the middle. Most importantly for interpretations with respect to Indo-European languages, the dotted lines that delineate this cline in Figure 1 are arbitrary in including one CHG sample but not the other, and thus also excluding the Neolithic Iran samples, even though they are just next to it (see figure reproduced and annotated below). The paper also makes it all the clearer that core Yamna was essentially an incoming population: 80% of its ancestry originated further south, and most of that ultimately from the Caucasus/Zagros region.

That is of course where other hypotheses on Indo-European origins had long placed the family's homeland, whether on linguistic grounds (Gamkrelidze & Ivanov 1984, 1995) or archaeological ones (Renfrew 1987). Now that this new paper supports that original homeland from genetic data too, the next big questions are obvious. Which branch(es) of the language family spread north to end up as core Yamna on the Steppe, and to emerge later from there? And which branches never went through the Steppe, but emerged independently out of the South Caucasus/Zagros homeland, spreading in directions other than northwards?

On those questions, this paper brings no new data, but just cites the same team's past claims. As in Lazaridis *et al.* (2022), this paper does now accept that the Anatolian branch did not emerge from the Steppe. This therefore contradicts the Steppe hypothesis, which had always specifically claimed that Anatolian also emerged from the Steppe: see arrow 1 on the map in Anthony & Ringe (2015). And the 'Steppe hypothesis' was of course named in the first place for where it set the family's original homeland.

Beyond conceding on the Anatolian branch, this new paper continues to maintain that all other branches emerged only from the Steppe. Indeed the authors would have the name 'Indo-European' refer now only to those, to exclude Anatolian – notwithstanding universal agreement that Anatolian belongs to the family. But Pandora's box is already opened. If one branch of the family certainly did not emerge from the Steppe, how solid really is the case that all others did? The authors appeal to Anatolian as a popular candidate, in linguistic analyses, for having been the first branch to separate from the rest of the family. Without linguist authors they actually oversimplify how much (dis)agreement and clarity there is on this, and on whether the family diverged in a clear-cut sequence of successive binary branches at all. They suppose a "much-earlier" split, but in fact there is no good case for a long period of separate development of an 'all except Anatolian' branch (for Anthony & Ringe 2015 themselves it could be as little as 300 years), and more for an early branching in multiple directions.

In any case, the paper here bases its interpretations essentially on ancient DNA, and other past publications by the same Harvard team. So it is worth a recap of those papers, through the story of the Harvard team's apparent drive to confirm, through ancient DNA, the Steppe hypothesis as expounded particularly by their co-author and Steppe archaeologist David Anthony (e.g. Anthony 2007). He had hitherto insisted, as late as Anthony & Ringe (2015), that the Steppe hypothesis did not entail major demographic movements. Nonetheless, when Haak *et al.* (2015) found "massive migrations" into much of Europe from the Yamna culture on the Steppe, from c. 4800 BP, it was this demographic logic that most observers saw as making a strong case for this bringing Indo-European languages, too, into Corded Ware Europe. That made a good potential fit with several branches of Indo-European, but in fact not even all of the family's branches in Europe. So the search was then on to tick off the same Steppe ancestry for all remaining branches of Indo-European, too, by analysing ancient DNA from past populations that most plausibly would have spoken them.

Over the ensuing years, more ancient DNA papers duly reported on these other branches, including key papers by the Harvard team, and which they cite here to assert that all branches except Anatolian emerged from the Steppe. What they do not state, however, is just how low are the proportions of Steppe ancestry in these cases. The "massive

migrations” into Corded Ware northern Europe proved elusive elsewhere. For Lazaridis *et al.* (2022), “Anatolia is remarkable for its lack of steppe ancestry down to the Bronze Age” – that only appears remarkable, though, if one expected it in the first place, because one had presumed the Steppe hypothesis. This new paper now states that “The CLV people contributed ... at least one-tenth of the ancestry of Bronze Age central Anatolians, who spoke Hittite.” The CLV cline is not Yamna on the Steppe, anyway, and 10% is no “massive migration” that might be expected to rewrite the language identity of central Anatolia (and western Anatolia, where these languages also dominated). More likely it left no major linguistic effect; the Anatolian branch remains better explained by stronger candidates (see Lazaridis *et al.* 2022).

The willingness in the new paper to see even 10% ancestry as if a good case for a language spread is reminiscent of their interpretations in the cases of Mycenaean Greek (Lazaridis *et al.* 2016) and South Asia (Narasimhan *et al.* 2019). In both cases their own ancient DNA analyses support no “massive” migration of people of Steppe origin. On the contrary, overall percentages are generally very low, and in South Asia also too late for a plausible first arrival of Indic languages here (let alone Indo-Iranic as a whole). But however small and late, and however implausible that they replaced all languages from Iran right across to northern India, that is what has to be claimed for these weak signals, for the Steppe hypothesis to be right. Narasimhan *et al.* (2019) propose a scenario to try to make this work for modern Indic-speaking populations, but it cannot simultaneously work for different patterns in speakers of Iranian languages – whereas the early histories of these two sub-branches of Indo-European cannot be separated in this way, since both emerge from a common earlier Indo-Iranic branch.

The new paper thus makes various statements as if of fact, citing their own past papers, but which are entirely at variance with other perspectives, not least from archaeology. It opens with the claim that “people of the Yamnaya archaeological complex and their descendants ... transformed ... South Asia” (among other places, but again incongruously omitting Iran), for which the Steppe hypothesis has long imagined the ‘descendant’ culture responsible as Andronovo. But there is “absolutely NO archaeological evidence for any variant of the Andronovo culture either reaching or influencing the cultures of Iran or northern India in the second millennium. Not a single artifact of identifiable Andronovo type has been recovered from the Iranian Plateau, northern India, or Pakistan” (Lamberg-Karlovsky 2005: 155). There was no such “transformation” of South Asia by Yamnaya or their descendants.

In search of another candidate, Narasimhan *et al.* (2019) sought out samples from the Oxus (or ‘BMAC’) civilisation, but again effectively drew a blank: no significant Steppe ancestry, but largely Iranian continuity instead. What the main BMAC culture site at Gonur Depe does host, however, already by 4250 BP, is a burial of a horse and wagon with bronze wheel rims, and perhaps even *soma* (or *haoma*) preparation – supposedly good markers of early Indo-Iranic, for example, but genetically *not* from the Steppe. Indeed while their ultimate origins would lie in the Caucasus/Zagros homeland, there is no strong case for excluding that the long evolution into the distinctly Iranian branch proceeded in or near what is now Iran, and the distinctly Indic branch on the Indus, spreading then also along the Ganges.

Other than citing their own earlier papers, on other branches of Indo-European this paper leaves rather too much unsaid. They recognise the major ancestry component in the homeland region of “Neolithic people from [the] Zagros”, and in Figure 1b this component is hiding in plain sight: right alongside their CLV cline, next to its *source* end, and labelled ‘Iran’. Far more than any minor Steppe ancestry, too little too late, it is this ancestry component, from the Indo-European family’s Caucasus/Zagros homeland, that remains predominant in speakers of Iranian and Indic to this day (see Haak *et al.* 2015: Fig. S6). Narasimhan *et al.* (2019) seek to rule out that it simply spread eastwards bearing Indo-Iranic, but Maier *et al.* (2023) concede that they cannot; see also Broushaki *et al.* (2016).

Where this new paper leaves us is that the cat is out of the bag: not all Indo-European goes back to the Steppe. Now freed from the Steppe hypothesis as the base presumption (or aspiration) for interpreting aDNA data, we can look forward to a neutral re-evaluation of the most plausible candidates for tracing multiple other branches of the Indo-European family out of the original Caucasus/Zagros homeland, without all having to travel via the Steppe. Alongside Anatolian, these may include notably Greek, Armenian, Albanian and Indo-Iranic.

For fuller discussion of the genetic and archaeological contexts through which Indo-European language prehistory played out, see Heggarty *et al.* (2023: SM 13-19) in *Science*, also curiously not cited here. To download it and explore its Indo-European languages database, see <https://iecor.clld.org> and <https://paulheggarty.info/indoeuropean>. For updates to this commentary, see: <https://paulheggarty.info/blog>.

References

- Anthony, D.W. 2007. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton: Princeton University Press.
- Anthony, D.W., & Ringe, D. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1(1): p.199–219. <https://doi.org/10.1146/annurev-linguist-030514-124812>
- Braidwood, R. J. (1960). The Agricultural Revolution. *Scientific American* 203(3), 130–152. <https://www.jstor.org/stable/24940620>
- Broushaki, F., Thomas, M.G., Link, V., López, S., et al. 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353(6298): p.499–503. <http://doi.org/10.1126/science.aaf7943>
- Gamkrelidze, T.V., & Ivanov, V.V. 1984. *Indoevropskij jazyk i indoevropejcy: Rekonstrukcija i istoriko-tipologičeskij analiz prajazyka i protokultury. [The Indo-European language and the Indo-Europeans: A Reconstruction and Historical-Typological Analysis of a Proto-Language and a Proto-Culture]*. Tbilisi: Tbilisi University Press.
- Gamkrelidze, T.V., & Ivanov, V.V. 1995. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture*. Berlin: Mouton de Gruyter.
- Ghalichi, A., Reinhold, S., Rohrlach, A.B., Kalmykov, A.A., et al. 2024. The rise and transformation of Bronze Age pastoralists in the Caucasus. *Nature* 635(8040): p.917–925. <https://doi.org/10.1038/s41586-024-08113-5>
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555): p.207–211. <http://doi.org/10.1038/nature14317>
- Heggarty, P., Anderson, C., Scarborough, M., King, B., et al. 2023. Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* 381(6656): p.414, eabg0818. <https://doi.org/10.1126/science.abg0818>
- Lamberg-Karlovsky, C.C. 2005. Archaeology and language: the case of the Bronze Age Indo-Iranians. In E. Bryant & L. L. Patton (eds) *The Indo-Aryan Controversy: Evidence and Inference in Indian History*, 142–177. London: Routledge.
- Lazaridis, I., Patterson, N., Anthony, D., Vyazov, L., et al. 2025. The genetic origin of the Indo-Europeans. *Nature*, 1–11. <https://doi.org/10.1038/s41586-024-08531-5>
- Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açikkol, A., et al. 2022. The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* 377(6609): p.eabm4247. <https://doi.org/10.1126/science.abm4247>
- Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., et al. 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature* 536(7617): p.419–424. <http://doi.org/10.1038/nature19310>
- Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., et al. 2023. On the limits of fitting complex models of population history to f-statistics M. Nordborg, M. Przeworski, D. Balding, & C. Wiuf (eds). *eLife* 12: p.e85492. <https://doi.org/10.7554/eLife.85492>
- Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., et al. 2019. The formation of human populations in South and Central Asia. *Science* 365(6457): p.eaat7487. <http://doi.org/10.1126/science.aat7487>
- Renfrew, C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.