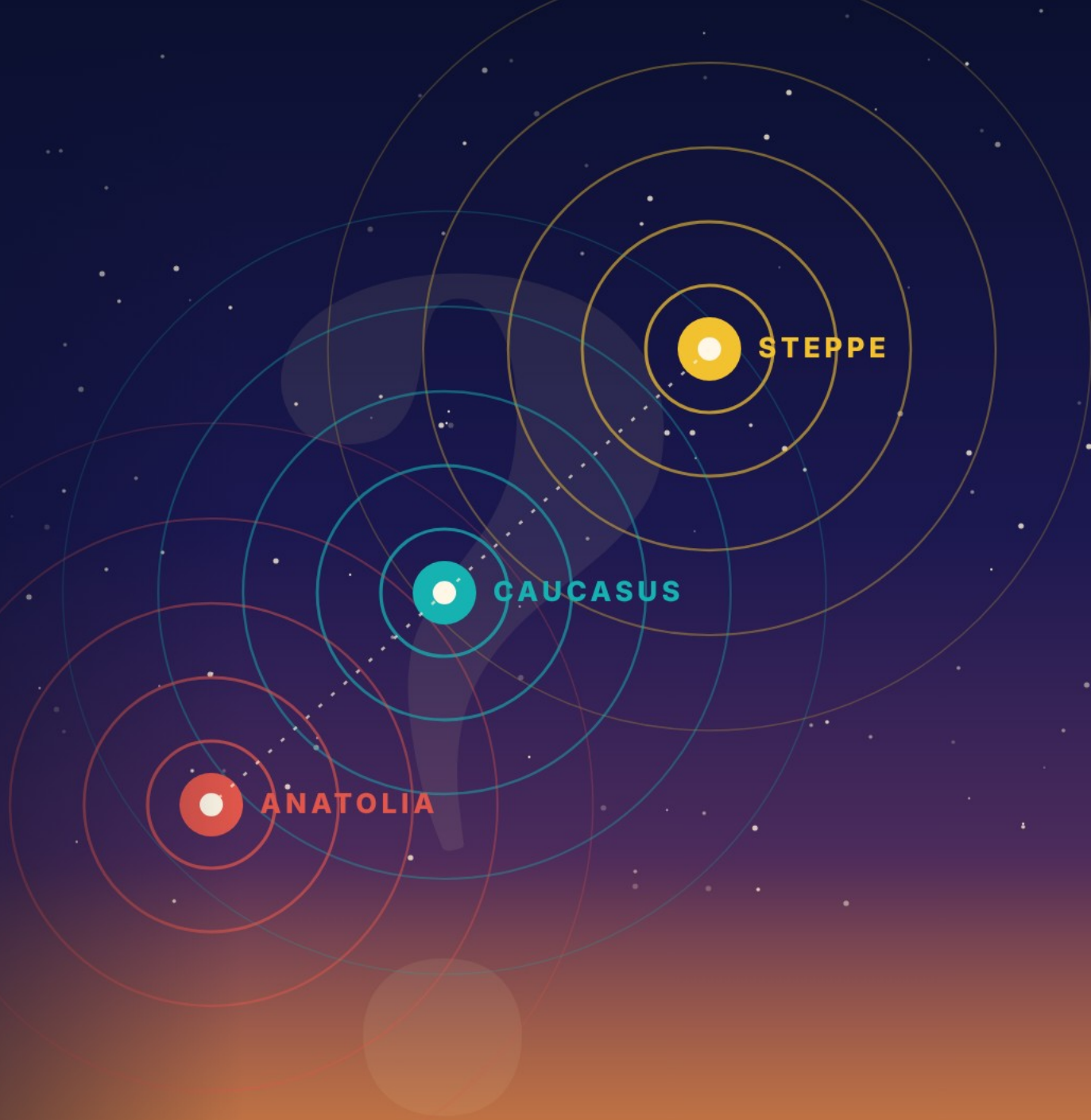
INDO-EUROPEAN COURSE · MODULE 1

01

Where Did the Indo-European Languages Begin?

Steppe, Anatolia or the Caucasus: what genes, languages and archaeology say



YSATIS

@airyanematlas · airyanematlas.github.io

What this module covers

1. **The puzzle**
   Why the origin of the family is still debated

2. **Three kinds of evidence**
   Linguistics, archaeology, ancient DNA

3. **Three models**
   Steppe, Anatolian farming, hybrid

4. **What ancient DNA found**
   Yamnaya, Anatolia and the Caucasus-Lower Volga people

5. **What linguistic trees say**
   Heggarty et al. 2023 and the dating dispute

6. **Consensus and open questions**
   What is settled and what is not

PART 1

01

Evidence and models

What kinds of evidence there are, and the three answers



A 200-year-old puzzle

Almost half

of the world's population speaks an Indo-European language

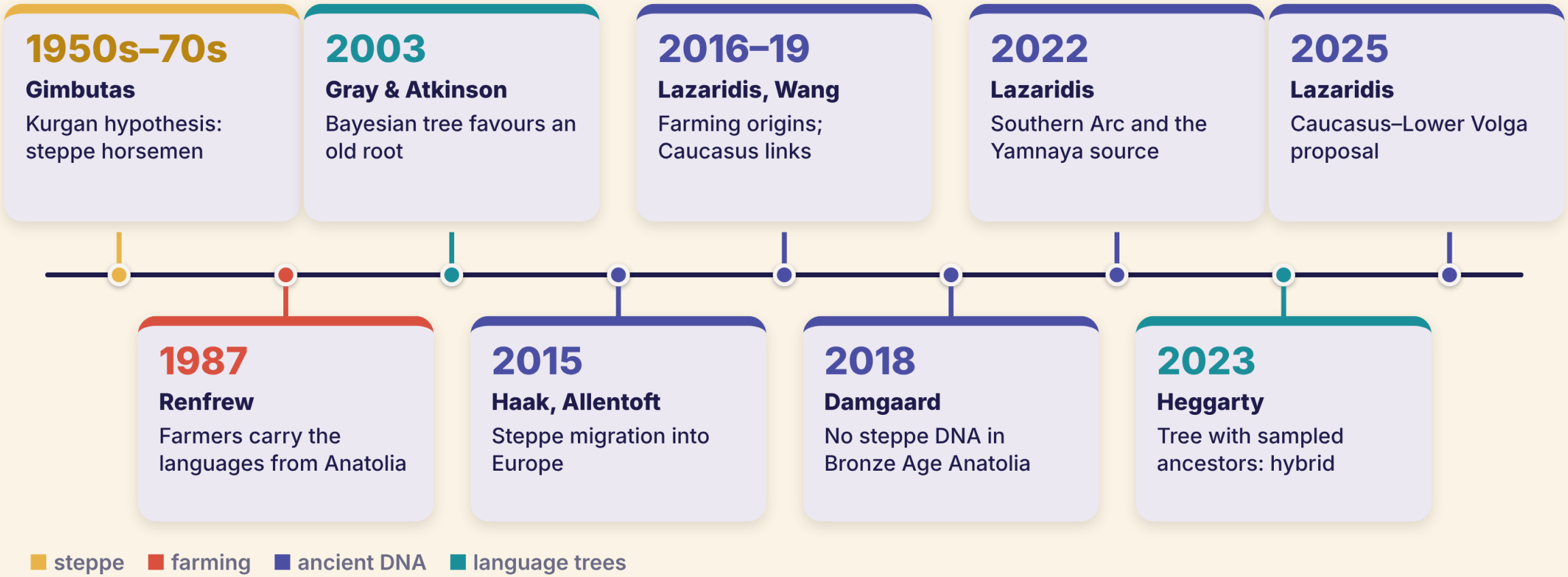
Source: Heggarty et al. 2023, Science

- The family runs from Icelandic to Bengali; its ancestor was never written down.
- Scholars reconstruct it by comparing languages.
- Where, when and why it spread is still disputed.



Map: nerdy.maps, CC BY-SA 4.0 (Wikimedia Commons)

Eighty years of the homeland debate



Tier A/B: dates from the papers cited in this module; Gimbutas and Gray and Atkinson 2003: standard history (verify). Heggarty 2023 and Lazaridis 2025 as cited in slides 17 and 21.

Three kinds of evidence



Linguistics

Compare vocabulary and grammar across languages; build family trees and estimate divergence dates with Bayesian methods.



Archaeology

Track material cultures such as Yamnaya: burials, settlements, wagons and horses, and how they spread.



Ancient DNA

Trace ancestry and migration from hundreds of ancient genomes, including a 2025 study of 428 individuals.

DNA shows how people moved, not which language they spoke; that gap is why the debate continues.

Three models of origin

Steppe

Proto-Indo-European spoken by herders on the Pontic-Caspian steppe; Yamnaya expansion from about 3300 BCE.

Advocate: David Anthony

Anatolian farming

The family spread from Anatolia with agriculture, more than 9,000 years ago.

Advocate: Colin Renfrew (1987)

Hybrid

A root south of the Caucasus, with a branch moving north onto the steppe; Anatolian and Yamnaya share West Asian ancestry.

Heggarty et al. 2023; Lazaridis et al. 2025

Three models side by side

	Steppe	Anatolian farming	Hybrid / Caucasus
Where	Pontic-Caspian steppe	Anatolia	South of the Caucasus; North Caucasus–Lower Volga
When	Yamnaya from c. 3300 BCE	More than 9,000 years ago	Proto-Indo-Anatolian c. 4400–4000 BCE
How	Pastoralists expand; language spreads through clans and clients	Farmers spread about 1 km a year (demic diffusion)	A root in the south, with one branch moving north onto the steppe
Strongest evidence	Steppe ancestry in Corded Ware: about 75%	Dated trees: root about 8,120 years ago	About 4/5 of Yamnaya ancestry from CLV people
Hardest question	Anatolian carries no steppe ancestry	Later DNA shifts come from the steppe	Does it still derive other branches from the steppe?
Associated with	Mallory, Anthony	Renfrew; Gray and Atkinson	Heggarty 2023; Lazaridis 2025

Tier B: synthesis of slides 5 to 21 and Pereltsvaig and Lewis 2015 (pp. 40-45), Heggarty et al. 2023, Lazaridis et al. 2025. 'Hardest question' cells are the module author's summary, not quotations.

The steppe hypothesis

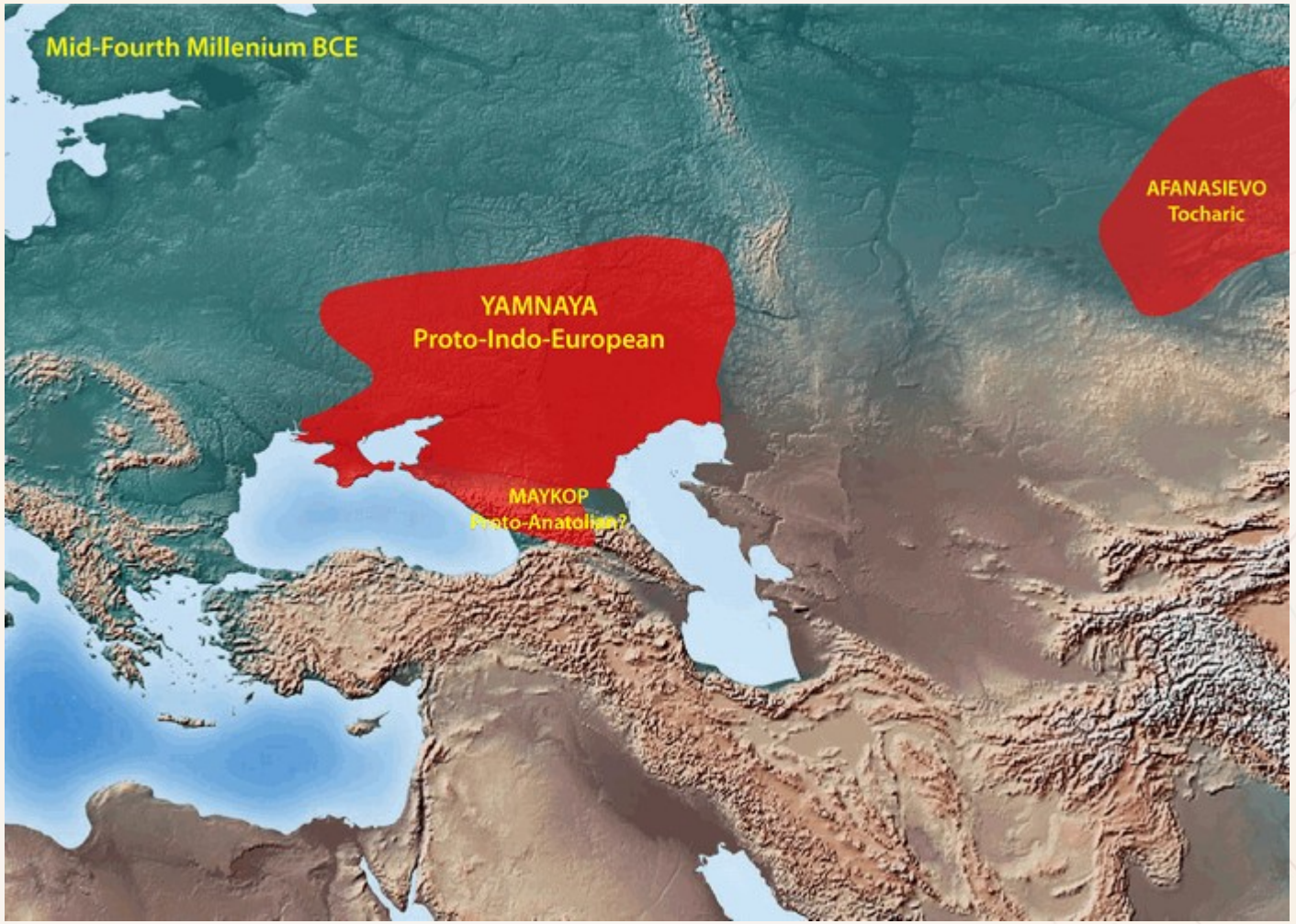
- The Yamnaya complex appears around 3300 BCE on the steppes north of the Black and Caspian Seas.
- By about 3000 BCE it spans from Hungary in the west to Kazakhstan in the east.
- Between 3300 and 1500 BCE, Yamnaya people and their descendants are proposed to have spread Indo-European languages across Eurasia.
- Anthony ties the story to horse domestication and the mobility it allowed.



Yamnaya kurgan, Hungary. Cseh, Molnár, Makó & Sümegi, CC0 (Wikimedia Commons)

Two terms
Yamnaya: Russian for "pit-grave", from the burial practice.
Kurgan: a burial mound (tumulus).

The steppe world, mid-4th millennium BCE



Map: Kees Huyser, Livius.org, CC0

- Yamnaya is labelled Proto-Indo-European; far to the east, Afanasievo is labelled Tocharic.
- Maykop, just south of the steppe in the Caucasus, is tagged "Proto-Anatolian?", the map-maker's own question mark.
- These labels are one interpretation. Compare them with the 2025 genetic model on later slides.

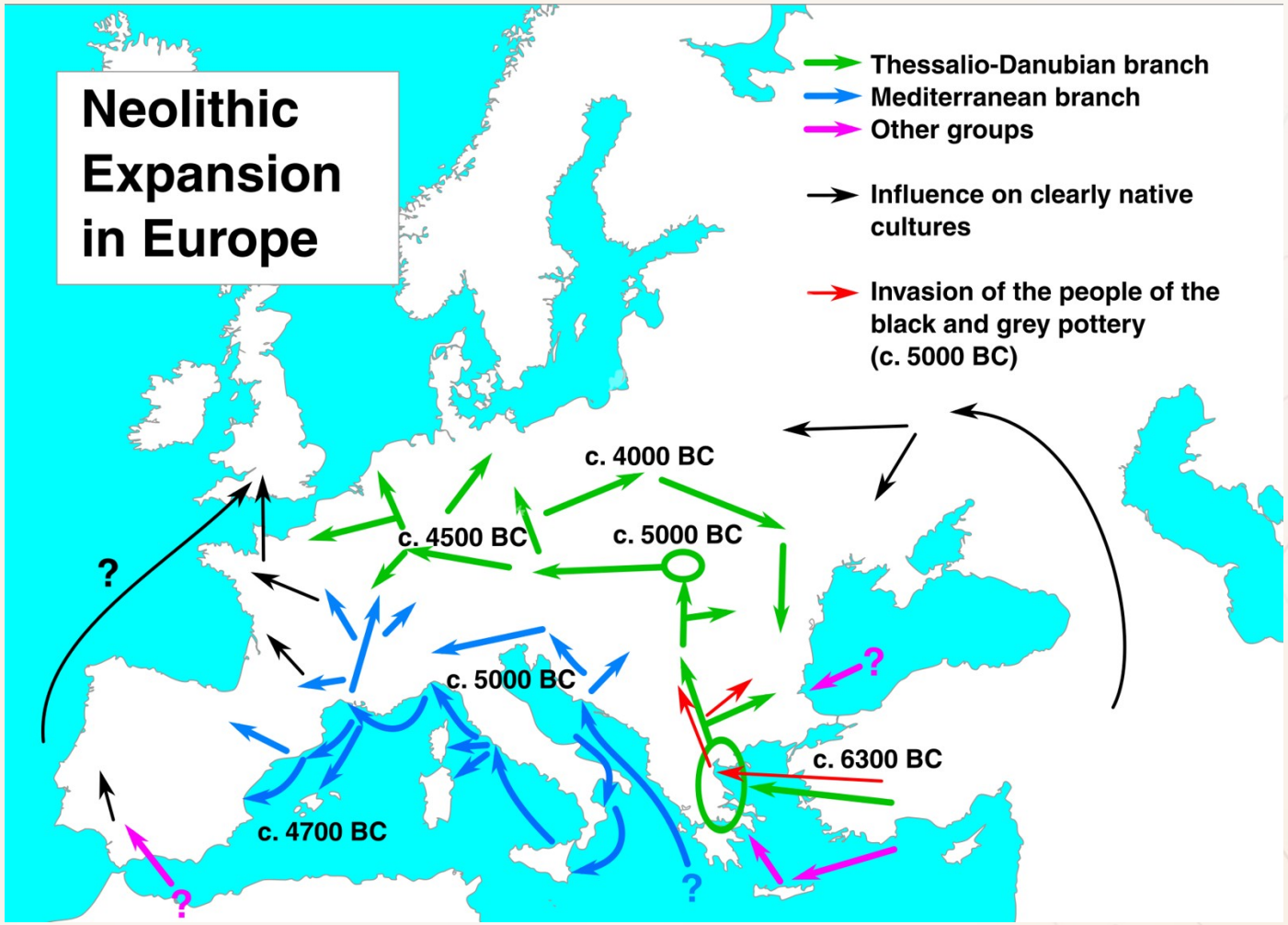
The Anatolian farming hypothesis



Çatalhöyük excavation. Volker Höhfeld, CC BY-SA 4.0 (Wikimedia Commons)

- Renfrew (1987): one event could explain the origin of the family, the spread of farmers from Anatolia after about nine thousand years ago.
- Hittite and other Anatolian languages appear in tablets about 4,000 years old and are deeply divergent from their relatives.
- Ancient DNA weakened the case: Anatolian samples show little or no steppe-like ancestry, and the Anatolian languages lack the full wagon and wheel vocabulary found elsewhere.

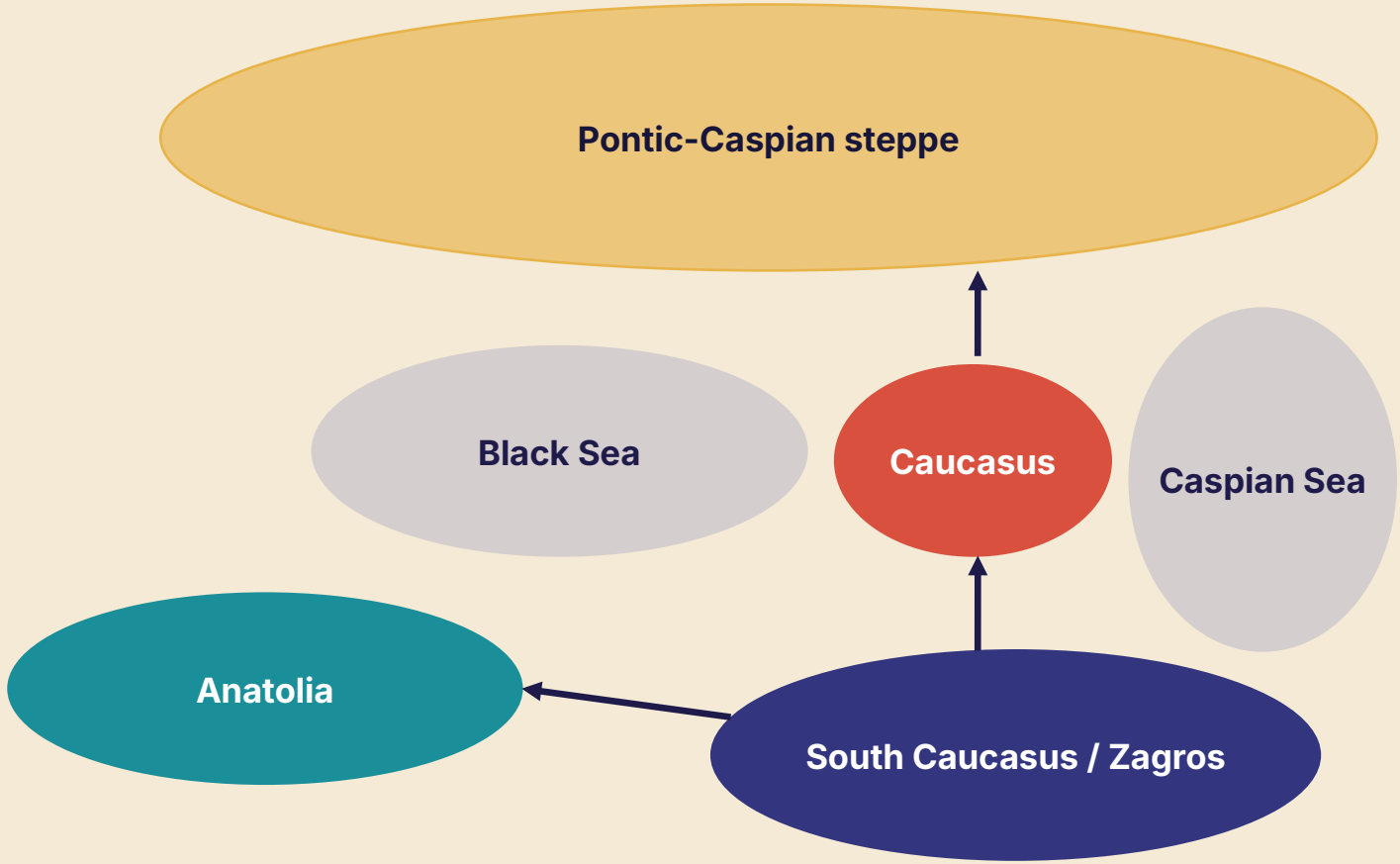
Farming spread in Europe



Map: Cthuljew, public domain (Wikimedia Commons)

- The map traces Neolithic farming across Europe from about 6300 BCE to 4000 BCE, in Thessalio-Danubian and Mediterranean branches.
- This is the kind of farmer dispersal Renfrew linked to the spread of Indo-European.
- It maps archaeological cultures, not languages; the language link is the contested part.

Candidate homelands (schematic map)



Schematic only: not to scale

Steppe model

Origin in the north; expansion from the Yamnaya

Farming model

Origin in Anatolia, spreading with farmers

Hybrid model

Roots south of the Caucasus; branches to Anatolia and the steppe

PART 2

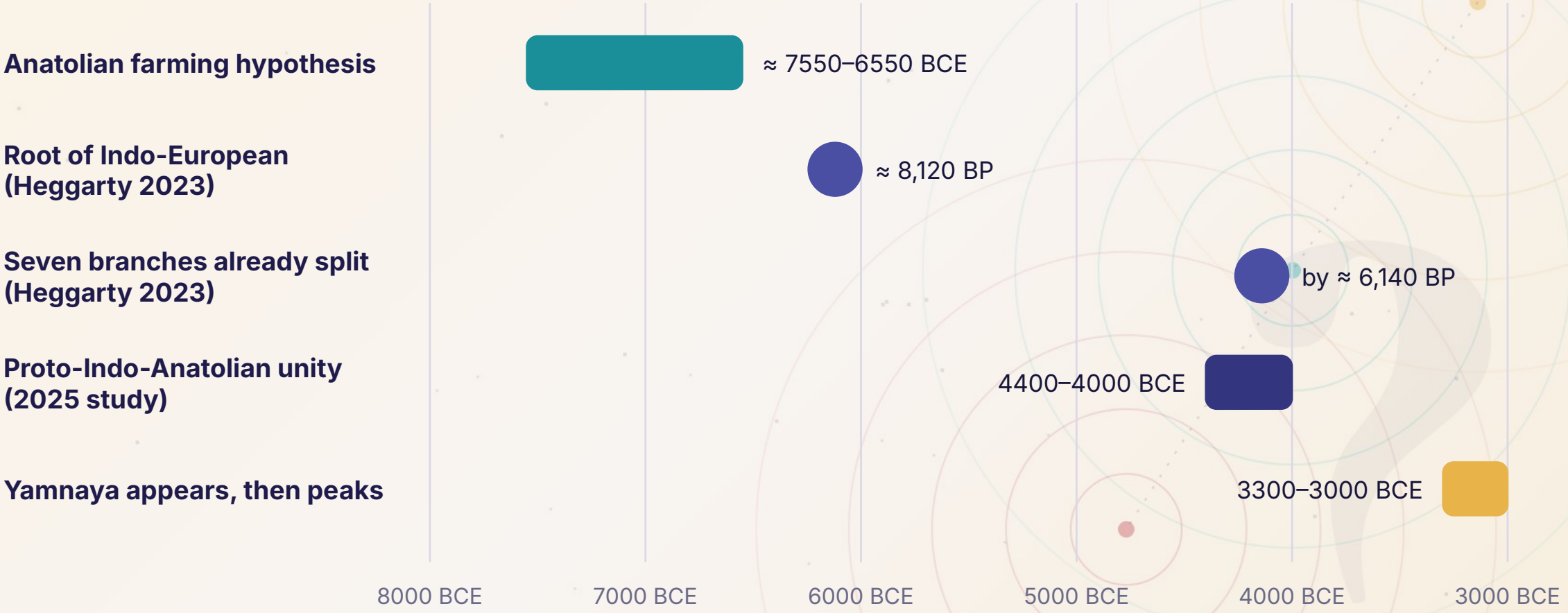
02

The steppe and farming

Maps, dates and two ways to frame the question



What each model implies about timing



Approximate dates. Heggarty et al. count BP from AD 2000, so 8,120 BP is about 6120 BCE and 6,140 BP about 4140 BCE. The Yamnaya bar shows appearance to maximum extent.

Two ways to frame the question

Renfrew 1987

Farmers spread from Anatolia, about 1 km a year (demic diffusion). Migration and word-based reasoning set aside.

Revised steppe

Pastoralist clans intrude into farming zones; language spreads through chiefs and clientage. Pontic steppe, horse (Mallory, Anthony).

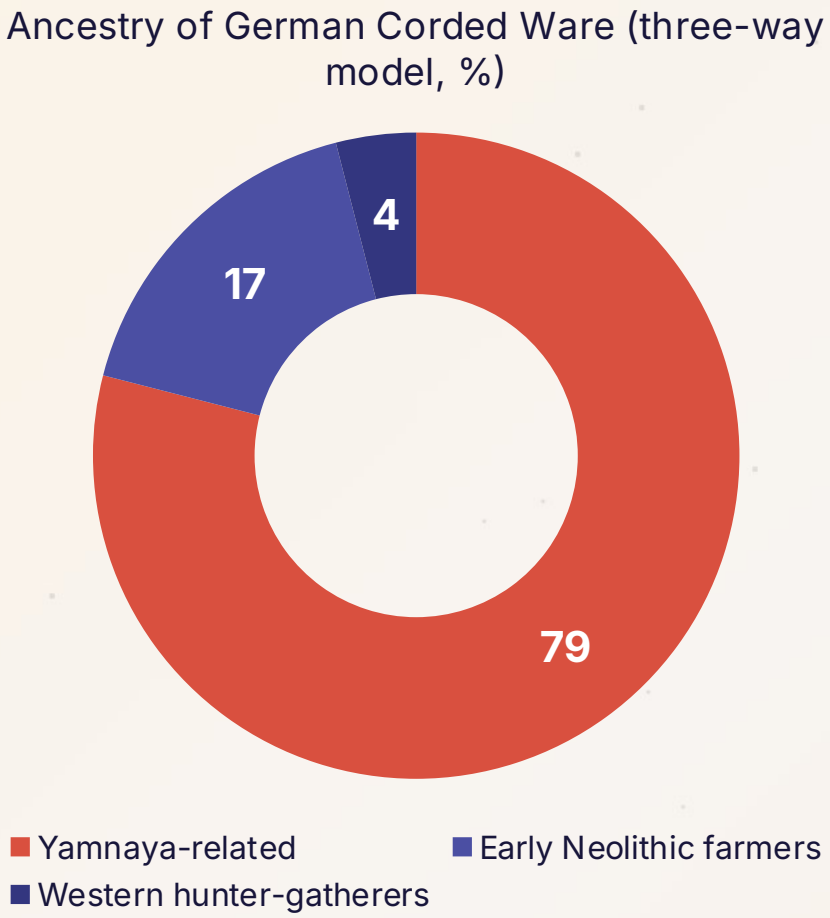
Options left out

Armenian (Gamkrelidze and Ivanov) and Balkan (Diakonoff) theories, ignored by the two-way test.

Models differ on mechanism as well as place: migration versus gradual diffusion.

Tier A for how positions are described; secondary source. Pereltsvaig and Lewis 2015, pp. 40–45, 49 (supplied; the authors argue against the Anatolian view).

Corded Ware: a steppe migration



73.1 ± 2.2%

Yamnaya-related ancestry (combined estimate)

≈ 4,500 BP

a "relatively sudden" migration into Central Europe

- About 75% of the ancestry of Corded Ware people from Germany traced to the Yamnaya.
- Genes show a migration, not the language spoken; the data cannot locate the homeland.

PART 3

03

What DNA and trees say

Corded Ware, Yamnaya, Anatolia and the Caucasus-Lower Volga link



Yamnaya: half steppe, half West Asian

≈ 1 : 1

eastern hunter-gatherer and Near Eastern-related ancestry

Haak et al. 2015

13–17%

Anatolian-farmer-related ancestry in Yamnaya groups (one model)

Wang et al. 2019

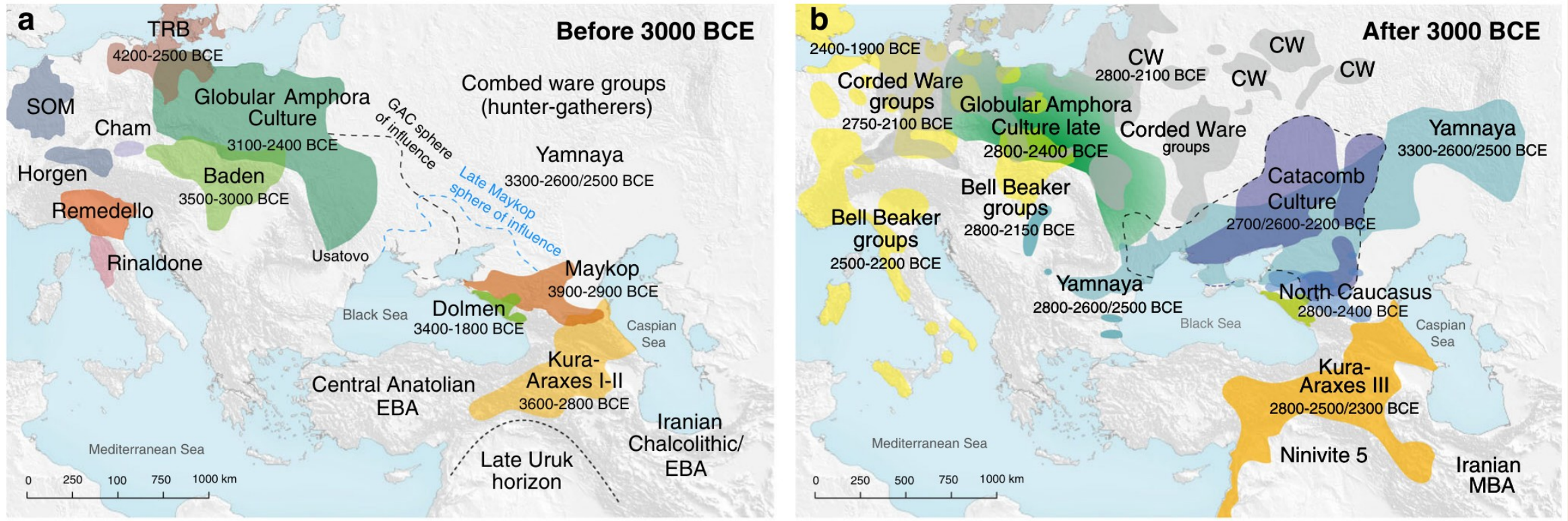
≈ 7,000 yr

ago: migration from West Asia into the steppe begins

Lazaridis et al. 2022

- The Yamnaya mixed northern (Eastern hunter-gatherer) and southern (Caucasus and West Asian) ancestry.
- Their expansion into the Caucasus was "a return to the homeland of about half their ancestors" (Lazaridis et al. 2022).

The Caucasus: a contact zone



Archaeological cultures before and after 3000 BCE. Wang et al. 2019, Nature Communications 10:590, Fig. 1a-b, CC BY 4.0 (cropped)

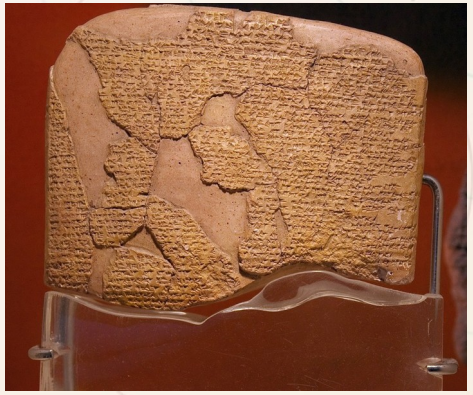
- Two genetic clusters, Steppe and Caucasus, stayed distinct for 3000 years, yet the mountains were "no barrier" to movement (Wang et al. 2019).
- Contact around the Maykop culture helped form steppe ancestry; later gene flow absorbed it into highland groups (Ghalichi et al. 2024).

Hittite without steppe DNA

Language	Hittite and its relatives are Indo-European, attested in tablets about 4,000 years old.
DNA	Ancient Anatolian genomes show little or no Yamnaya-like steppe ancestry.
Vocabulary	Anatolian languages lack the full wagon and wheel vocabulary of the other branches.
Link	Yamnaya and Anatolian speakers share ancestry in West Asia (Caucasus and Zagros highlands).



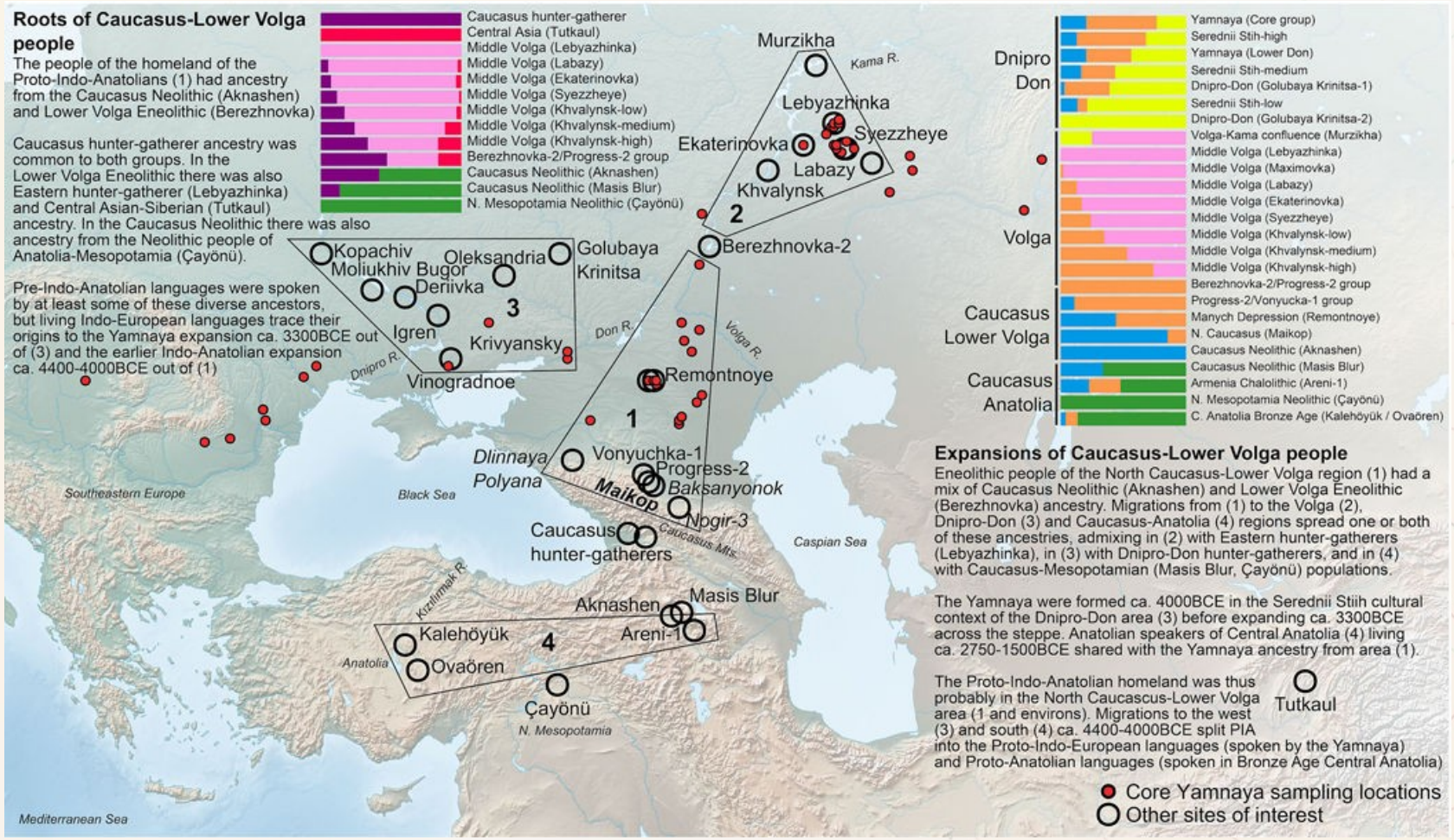
King's Gate, Hattusa. Carole Raddato, CC BY-NC-SA (WorldHistory.org)



Tablet of the Treaty of Kadesh, a Hittite-Egyptian peace settlement (Istanbul Archaeology Museum).

Photo: Iocanus, CC BY 3.0 (Wikimedia Commons)

A 2025 proposal for the homeland



Lazaridis et al. 2025, Nature 639:132-142, Extended Data Fig. 5, CC BY 4.0 (unaltered)

- The authors place Proto-Indo-Anatolian in the North Caucasus–Lower Volga area (1 on the map).
- Moves west (3) and south (4), about 4400–4000 BCE, split Indo-European from Anatolian.
- Caution: genetics cannot show who first spoke these languages.

The Caucasus-Lower Volga link

428
ancient individuals sampled in the 2025 study

≈ 4/5
of Yamnaya ancestry from CLV people

≥ 1/10
CLV ancestry in Bronze Age Central Anatolia

4400–4000
BCE: proposed unity of Proto-Indo-Anatolian speakers

- Yamnaya ancestors formed around 4000 BCE when CLV people moved west and mixed with Ukrainian hunter-gatherers along the Dnipro.
- The same CLV people also entered Anatolia, likely from the east, tying the Anatolian and steppe branches to a shared source.

The numbers behind the debate

428
ancient individuals, 2025 study

73.1%
Yamnaya-related ancestry in Corded Ware (± 2.2)

≈4/5
of Yamnaya ancestry from CLV people

≥1/10
CLV ancestry in Bronze Age Central Anatolia

8,120
years BP: median root of the language tree

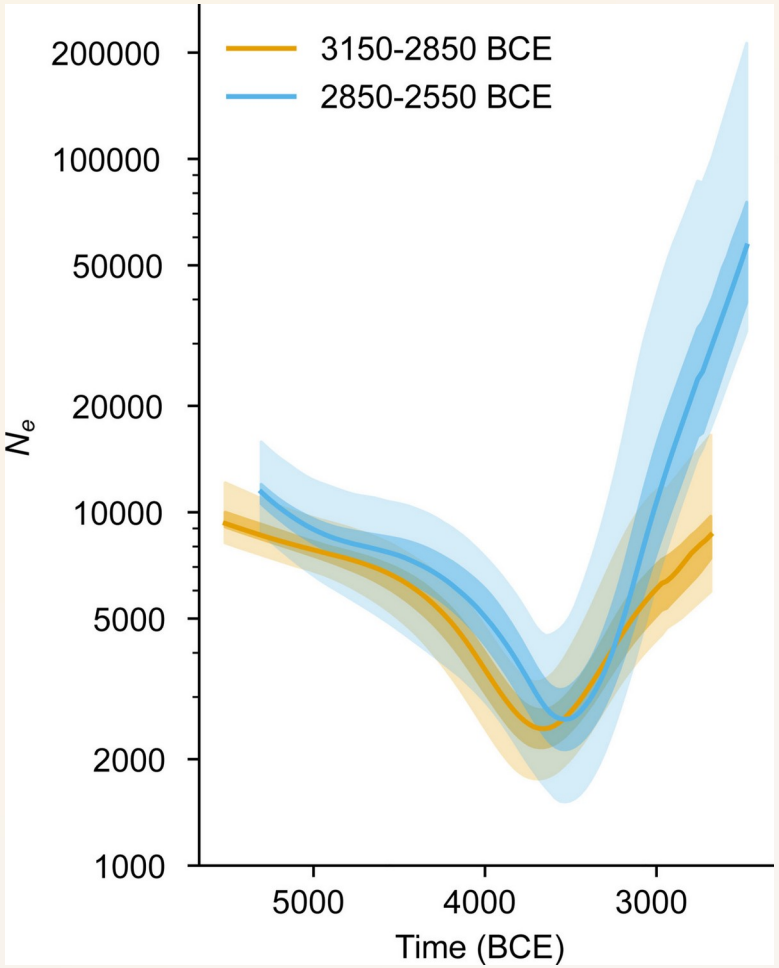
6,140
years BP: already split into seven branches

4400–4000
BCE: proposed Proto-Indo-Anatolian unity

3300
BCE: Yamnaya appears; widest by about 3000 BCE

Tier A: figures as stated on slides 13, 17, 18 and 21 (Lazaridis et al. 2025; Heggarty et al. 2023; Haak et al. 2015).

The Yamnaya population explosion



Lazaridis et al. 2025, Nature 639:132-142, Fig. 4, CC BY 4.0 (unaltered)

- Effective population size of Yamnaya ancestors rose from a few thousand to about an order of magnitude more.
- The two sampling windows (3150–2850 and 2850–2550 BCE) agree on an expansion at 3642–3374 BCE.
- Yamnaya ancestors formed about 4000 BCE; the archaeological culture appears about 3300 BCE, after the growth began.

Asia: a later, weaker steppe signal

Anatolia, Central and South Asia

Early Yamnaya-related migrations had limited direct genetic impact; no steppe ancestry in Bronze Age Anatolia.

Damgaard et al. 2018

South Asia

Steppe ancestry integrates in the Swat Valley in the first half of the 2nd millennium BCE; up to 30% of ancestry in modern South Asian groups.

Narasimhan et al. 2019

Bactria–Margiana (BMAC)

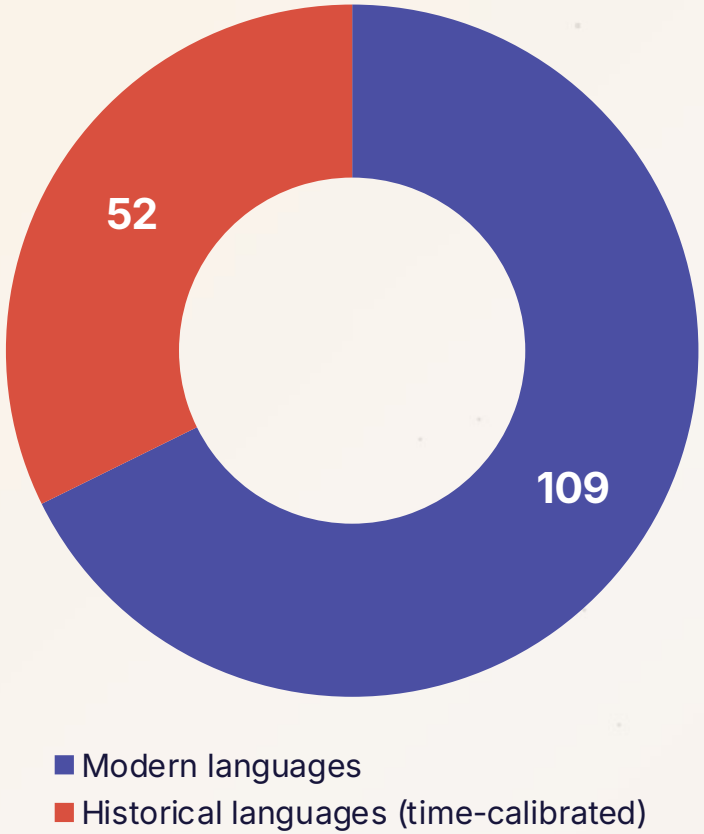
The main BMAC population carried no steppe ancestry; outliers appear around 2100–1700 BCE.

Narasimhan et al. 2019

Contested: Heggarty (2025) argues these signals are too small and late to explain the spread of Indo-Iranic.

What language trees say

Heggarty et al. 2023 dataset (161 languages)



≈ 8,120

years BP: median root age

≈ 6,140

years BP: Indo-European already split into seven branches

- Bayesian trees with "sampled ancestors" suggest few ancient languages are direct ancestors of later ones.
- The authors read the result as a hybrid of the farming and steppe hypotheses: an origin south of the Caucasus, with a branch going north.

A critic of the 2012 language tree

283

empty columns found in the 2012 data matrix (admitted 2013)

≈500

years: the shift in the revised tree's main divergence

3,500 vs 1000 CE

Romani split in the tree, versus the historical record (their test case)

- A contested, polemical voice (tier C): it criticises one research programme, and the authors say their issue is method, not that Anatolia is impossible (p. 51).
- Their own view: they tentatively endorse the steppe hypothesis (p. 49).
- Read with the newer language-tree slides (21 and 24).

Tier C. Pereltsvaig and Lewis 2015, pp. 1–3, 9–10, 49, 51, 58–60 (supplied, partial). Rebuttals and revisions by Gray and Atkinson are not in the supplied text.

PART 4

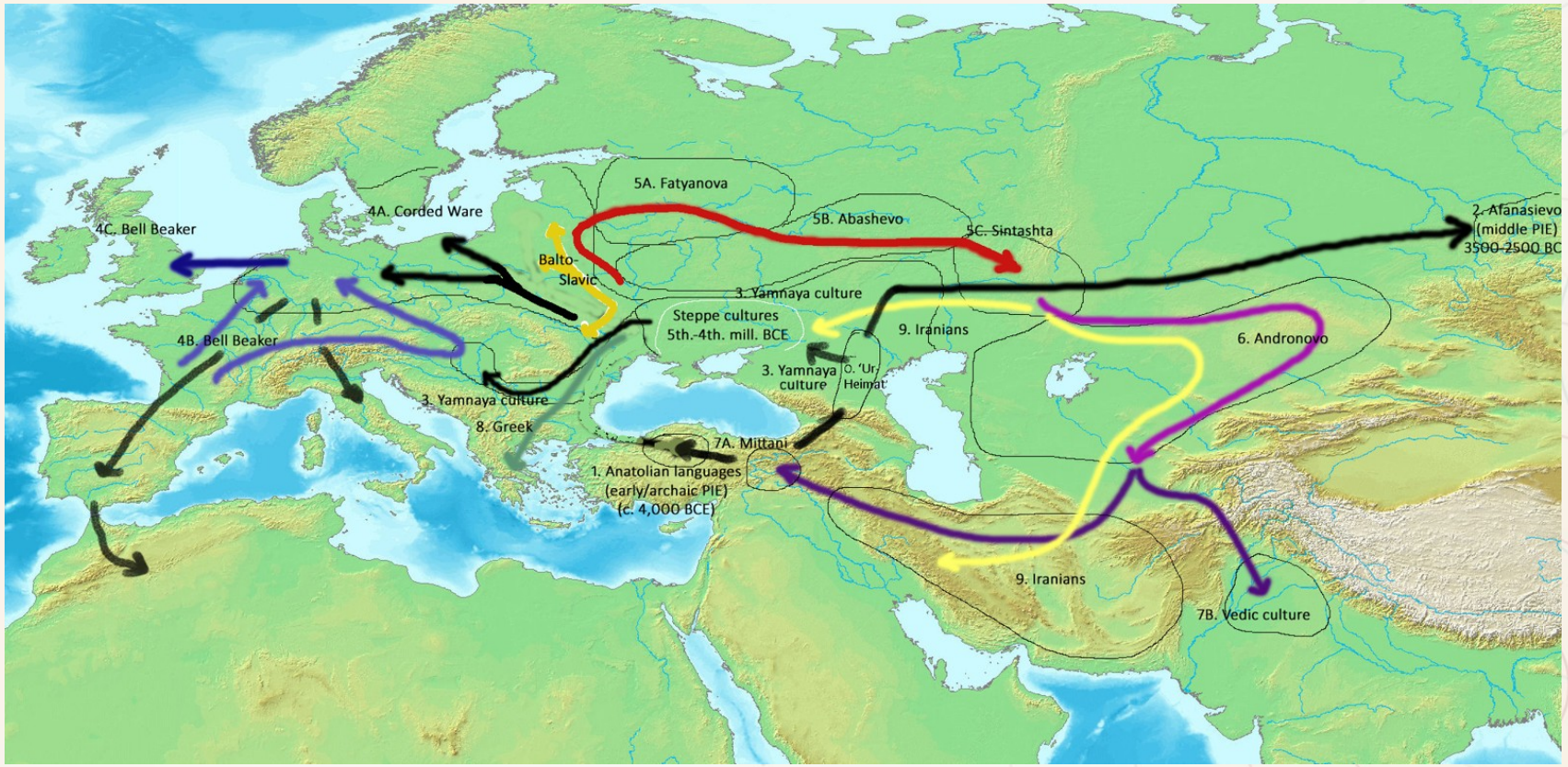
04

Debates and open questions

Where scholars disagree, and what is settled



An older overview of the expansions



Map: Joshua Jonathan, CC BY-SA 4.0 (Wikimedia Commons)

- A classic steppe-model overview: arrows spread from the steppe to Europe, Anatolia, Iran and India.
- Routes and dates are schematic, and ancient DNA has since refined them.
- Use it for orientation, not proof; Module 2 revisits each culture.

Two hypotheses for Proto-Indo-Anatolian

Hypothesis A

Proto-Indo-Anatolian was spoken by a population with high Eastern hunter-gatherer ancestry that influenced Anatolia linguistically but left little ancestry.

Problem: how could a group with little genetic impact change the language?

Hypothesis B

Proto-Indo-Anatolian was spoken in West Asia and the Caucasus by people with low or no Eastern hunter-gatherer ancestry, affecting both Anatolia and the steppe.

Fits the mix of Anatolian, Hattic and Hurrian speakers in Bronze Age Anatolia.

The authors of the 2022 study favour a West Asian progenitor; sampling gaps remain.

Where scholars disagree

1	**Is the dated tree reliable?**	Kassian and Starostin, in an informal review, argue the Heggarty tree lacks statistical robustness and contradicts established knowledge.
2	**Did Anatolian come from the steppe?**	Heggarty argues that the 2025 genetic paper now accepts Anatolian did not emerge from the steppe, yet still derives every other branch from it.
3	**Was there a Yamnaya "invasion"?**	Some critics point to discrepancies between archaeological, linguistic and genetic data in the mass-migration scenario.

Six facts every homeland model must explain

Language

Anatolian is the first, or among the first, branches to split

Heggarty 2023

Vocabulary

Anatolian lacks the full wagon and wheel vocabulary of other branches

Anatolian vs. other branches

Genes: the source

Yamnaya = steppe hunter-gatherers + Caucasus / West Asian ancestry

Haak 2015; Lazaridis 2025

Genes: Anatolia

Little or no Yamnaya-like steppe ancestry in Bronze Age Anatolia

Damgaard 2018

Archaeology

Yamnaya spreads fast from about 3300 BCE; widest by 3000 BCE

Yamnaya complex

Dates

Roots near 4400 BCE (genes) versus about 6100 BCE (language trees)

2025 vs 2023

Tier A/C: facts as stated on slides 6, 16, 17, 21 and 26; the grouping into six is a teaching synthesis.

Who's who in the debate

Marija Gimbutas

Kurgan hypothesis: steppe pastoralists as the Indo-Europeans

Colin Renfrew

1987: language spread with Neolithic farmers from Anatolia

David Anthony

Steppe, horse and wagon; Anthony and Ringe 2015

J. P. Mallory

Steppe archaeology and language; Mallory and Adams 2006

Paul Heggarty et al.

2023 Bayesian tree with sampled ancestors: a hybrid model

Lazaridis, Reich et al.

2025: Caucasus–Lower Volga source for Yamnaya and Anatolian

Kassian & Starostin

2023 informal review: the dated tree lacks robustness

Gamkrelidze & Ivanov

Armenian-highlands theory, left out of the two-way test

Igor Diakonoff

Balkan theory, also left out of the two-way test

Tier A/B: roles as described in slides 12, 17, 21, 25 and in Pereltsvaig and Lewis 2015; Gimbutas, Anthony and Mallory: standard (verify in Mallory and Adams 2006).

Three misreadings to avoid

THE MISREADING

"The genes show where the language came from."

THE BETTER READING

Genes show where ancestry came from. Language can move without large migrations, and migrations can happen without language change. Treat DNA as one line of evidence next to language and archaeology.

THE MISREADING

"There is a settled answer."

THE BETTER READING

Even the best papers disagree on root date, homeland and the route of the later branches. Present any single homeland map as provisional.

THE MISREADING

"Anatolian is a daughter of the steppe."

THE BETTER READING

The 2025 paper places Anatolian and the steppe branches as sisters with a shared West Asian source; critics dispute how fully. Ask which paper and which model each claim comes from.

Synthesis of slides 13, 16, 17, 24 and 25.

Five questions for any “origin of Indo-European” headline

1. **Which model?**
   Steppe, Anatolian farming, or the hybrid?

2. **Which evidence?**
   Words, material culture or DNA, and how many individuals?

3. **Which date?**
   A tree date, a radiocarbon date or an ancestry date?

4. **Which branch?**
   Does the claim cover Anatolian, or only the later branches?

5. **Which source?**
   A peer-reviewed paper, a review or a commentary?

Synthesis by Claude from slides 3 to 26.

Settled and still open

Broadly supported

- Yamnaya expanded rapidly from about 3300 BCE and reached its widest extent by about 3000 BCE.
- Anatolian is the first, or among the first, branches to split (the order of the deepest branches is weakly supported) and does not carry steppe ancestry.
- Yamnaya and Anatolian speakers share ancestry rooted in West Asia.

Still contested

- Exact homeland: steppe, Caucasus-Lower Volga, or south of the Caucasus.
- Root date: about 4400 BCE or deeper, near 6100 BCE.
- Whether other branches (Greek, Armenian, Indo-Iranian) passed through the steppe.
- How much mass migration versus gradual spread each region saw.

Check your understanding

1. What are the three kinds of evidence used to study the origin of the family?
2. Who proposed the Anatolian farming hypothesis, and in which year?
3. When does the Yamnaya complex appear on the steppe?
4. About what share of Corded Ware ancestry traces to the Yamnaya?
5. What do the letters CLV stand for, and why do they matter?
6. What does Hittite DNA show about steppe ancestry?
7. What median root age did the 2023 language tree give?
8. What place and date does the 2025 paper propose for Proto-Indo-Anatolian?
9. Why can genes alone not settle the language question?
10. Name two points that are still contested.

Questions drawn from slides 3 to 26.

Check your understanding: answers

1. **Three kinds**
   Linguistics, archaeology and ancient DNA.

2. **Renfrew, 1987**
   Language spread with Neolithic farmers from Anatolia.

3. **About 3300 BCE**
   Widest extent by about 3000 BCE.

4. **About 75%**
   73.1 ± 2.2% in the combined estimate (Haak 2015).

5. **Caucasus-Lower Volga**
   The source of about 4/5 of Yamnaya ancestry (Lazaridis 2025).

6. **Little or none**
   No steppe ancestry in Bronze Age Anatolia (Damgaard 2018).

7. **About 8,120 years BP**
   Heggarty et al. 2023; seven branches by 6,140 BP.

8. **North Caucasus-Lower Volga, 4400–4000 BCE**
   Lazaridis et al. 2025.

9. **Ancestry is not language**
   Languages can spread without gene flow and the reverse.

10. **Homeland, root date**
    Also whether other branches passed through the steppe.

As above.

Key terms

Yamnaya
Bronze Age steppe pastoralists, about 3300–2600 BCE, known for kurgan burials and wagons.

Kurgan
A burial mound, typical of the steppe.

Proto-Indo-European
The reconstructed ancestor of the family; never written down.

Proto-Indo-Anatolian
The common ancestor of Anatolian and the other branches (2025 proposal).

CLV
Caucasus-Lower Volga: the proposed source population in the 2025 study.

EHG
Eastern hunter-gatherer: northern ancestry in Yamnaya.

Demic diffusion
Spread of a population, as with farmers expanding from Anatolia.

Bayesian tree
A language tree built with statistical models and dated using known events.

Sampled ancestor
An attested language treated as a possible ancestor of later ones.

Admixture
Mixing of previously separate populations.

Corded Ware
Late Neolithic culture of north and central Europe, with strong Yamnaya ancestry.

Hybrid model
A homeland south of the Caucasus with a branch onto the steppe.

Definitions summarised from this module and standard handbooks.

Takeaways

1

The steppe is not the whole story

Anatolian and Yamnaya share a West Asian source.

2

Dates depend on the method

Genetics and language trees give different roots.

3

Debate is ongoing

Treat any single homeland map as provisional.

Next module: archaeology of the steppe (Yamnaya, Corded Ware, Sintashta)

Key literature

- Anthony, D. & Ringe, D. (2015). The Indo-European homeland from linguistic and archaeological perspectives. Annual Review of Linguistics 1: 199–219.
- Renfrew, C. (1987). Archaeology and Language: The Puzzle of Indo-European Origins. Cape.
- Haak, W. et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522: 207–211.
- Allentoft, M. et al. (2015). Population genomics of Bronze Age Eurasia. Nature 522: 167–172.
- Damgaard, P. de B. et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science 360: eaar7711.
- Narasimhan, V. et al. (2019). The formation of human populations in South and Central Asia. Science 365: eaat7487.
- Lazaridis, I. et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature 536: 419–424.
- Heggarty, P. et al. (2023). Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. Science 381: eabg0818.
- Lazaridis, I. et al. (2022). The genetic history of the Southern Arc. Science 377: eabm4247.
- Wang, C.-C. et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus. Nature Communications 10: 590.
- Ghalichi, A. et al. (2024). The rise and transformation of Bronze Age pastoralists in the Caucasus. Nature 635: 917–925.
- Lazaridis, I. et al. (2025). The genetic origin of the Indo-Europeans. Nature 639: 132–142.
- Kassian, A. & Starostin, G. (2023). Informal review of Heggarty et al. 2023; Heggarty, P. (2025). Commentary on Lazaridis et al. 2025.
- Pereltsvaig, A. and Lewis, M. W. (2015). The Indo-European Controversy. CUP (partly read; polemical).

Citations compiled from abstracts and summaries; verify volume, page and author details against the originals before citing.

Image credits

Slide	Image	Author	License
1	Steppe horses, Azov-Sivash reserve	YegorGeologist	CC BY-SA 4.0
3	Map of main Indo-European languages	nerdy.maps	CC BY-SA 4.0
6	Yamnaya kurgan, Hungary	Cseh, Molnár, Makó, Sümegi	CC0
7	Steppe world, mid-4th millennium BCE	Kees Huyser (Livius.org)	CC0
8	Çatalhöyük excavation	Volker Höhfeld	CC BY-SA 4.0
9	Neolithic expansion in Europe	Cthuljew	Public domain
14	Cultures before and after 3000 BCE (Fig. 1a-b)	Wang et al. 2019 (Nat. Commun.)	CC BY 4.0
15	King's Gate, Hattusa	Carole Raddato (WorldHistory.org)	CC BY-NC-SA
15	Treaty of Kadesh tablet	locanus	CC BY 3.0
16	Origin of Indo-Anatolian languages (ED Fig. 5)	Lazaridis et al. 2025 (Nature)	CC BY 4.0
18	Trajectory of the Yamnaya expansion (Fig. 4)	Lazaridis et al. 2025 (Nature)	CC BY 4.0
21	Indo-European expansions map	Joshua Jonathan	CC BY-SA 4.0

Images were cropped but not otherwise altered. Check each license before reuse outside personal educational use.