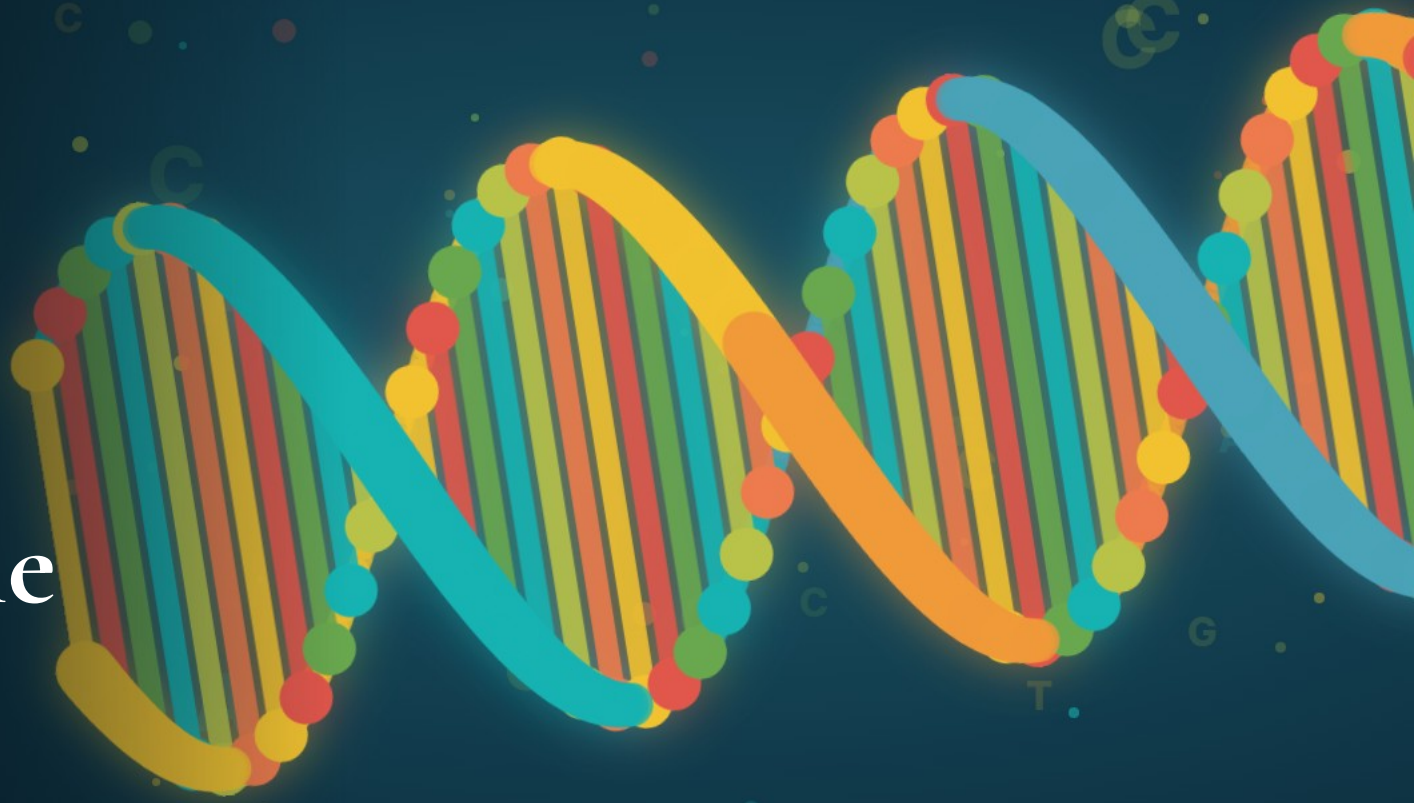
INDO-EUROPEAN COURSE · MODULE 3

03

Reading the Genes: Ancient DNA and the Indo-Europeans

Methods, ancestry streams, steppe migrations, and what DNA can and cannot say about language



YSATIS

@airyanematlas · airyanematlas.github.io

What this module covers

1 Reading the DNA

How samples become ancestry percentages, and the seven source populations.

2 Europe and the steppe

Haak, Allentoft and Lazaridis: who the Yamnaya were and what they did to Europe.

3 The Southern Arc

Anatolia, the Balkans and Armenia: where steppe ancestry did and did not arrive.

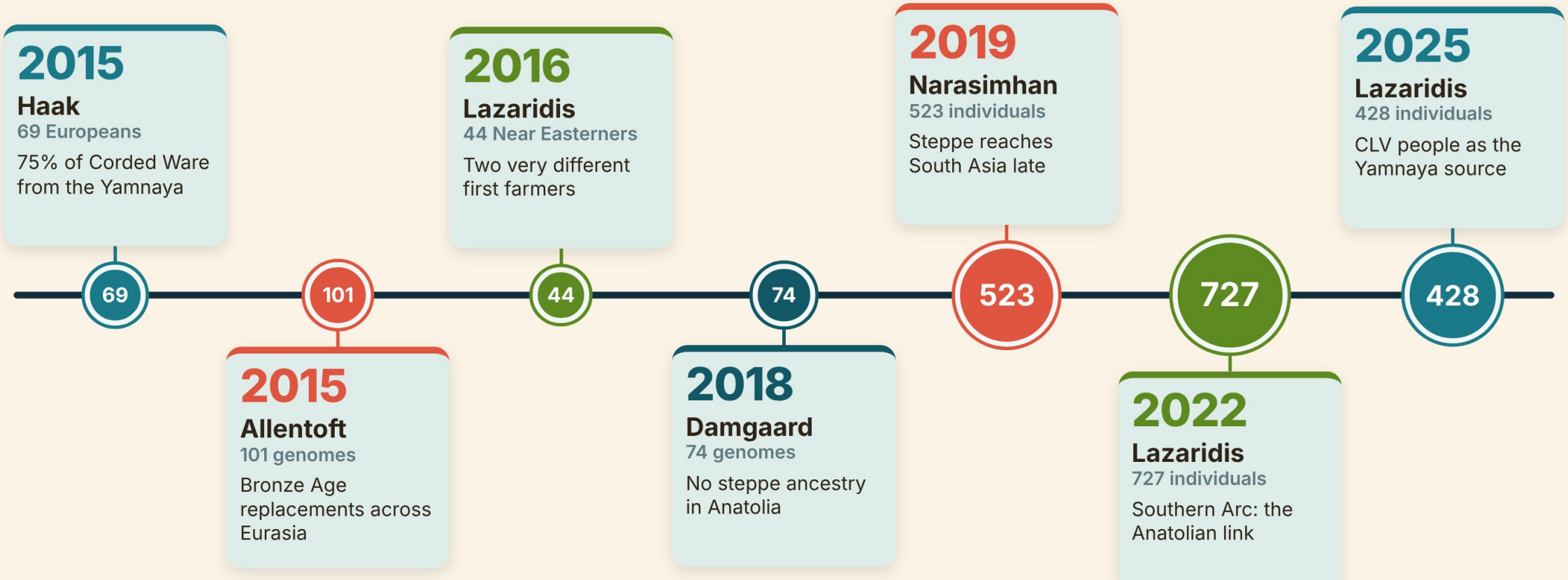
4 Asia, and the debate

BMAC, Swat, the ANI and ASI; then the limits of the method.

DNA tells us who moved and mixed. It does not tell us what they spoke.

Course plan. Archaeology in Module 2; languages in Module 4; Indo-Iranians in Module 5.

Ten years of ancient DNA



Circle size = number of individuals or genomes analysed.

Tier A: sample sizes from the abstracts of Haak 2015, Allentoft 2015, Lazaridis 2016, Damgaard 2018, Narasimhan 2019, Lazaridis 2022 and Lazaridis 2025.

Nine studies at a glance

Study	People	Region	Span	Headline
Lazaridis 2016	44	Near East	12,000–1,400 BC	Half of early Near Eastern ancestry from “Basal Eurasians”
Haak 2015	69 new	Europe, Russia	8,000–3,000 BP	Steppe migration into central Europe
Allentoft 2015	101	Eurasia	Bronze Age	Yamnaya spread; Corded Ware; Sintashta to Andronovo
Damgaard 2018	74	Inner Asia, Anatolia	9000–600 BC	Botai diverged; no steppe in Hittite-era Anatolia
Wang 2019	45	North Caucasus	3,000 years	North Caucasus split from the steppe; Maykop signals
Narasimhan 2019	523	South, Central Asia	8,000 years	Indus Periphery cline; steppe arrives late
Lazaridis 2022	727	Southern Arc	11,000 years	Yamnaya reach the Balkans; Armenia an enclave
Ghalichi 2024	131	Caucasus	6,000 years	Steppe ancestry forms; long Bronze Age stability
Lazaridis 2025	435	Steppe, Danube	6400–2000 BC	CLV source; Yamnaya formed about 4000 BC

Tier A: abstracts and methods of the nine papers. Spans and counts as printed; Damgaard span from the sample list (c. 9000 to 600 BC).

What ancient DNA is

Old, broken DNA

Bones and teeth hold short, damaged fragments, mostly from microbes. Warm places preserve it poorly.

Capture a panel

Enrich about 400,000 to 1.2 million known variable sites (SNPs). About 250 times less sequencing (Haak 2015).

Or read it all

Whole-genome sequencing: more data per person, much more cost (Allentoft 2015, Damgaard 2018).

Each person gives a noisy sample of their genome. Conclusions come from many people compared statistically.

Tier A: Haak et al. 2015, Nature 522:207–211 (abstract); Lazaridis et al. 2025, Nature 639:132–142; Allentoft et al. 2015; Damgaard et al. 2018.

PART 1

01

Reading the DNA

From samples to ancestry percentages



Four tools you will meet again and again

PCA

A map of similarity. People who share ancestry sit near each other; mixed groups fall between the sources.

qpAdm

Models a group as a mixture of source populations and tests which mixtures fit. Gives percentages and P values.

IBD segments

Long shared DNA segments reveal genealogical links, within and between cemeteries.

Dating

Admixture dates and population-size curves (HapNe-LD) give the when of mixing and growth.

A "fit" is a statement about the data, not proof that the model is the true history.

Tier A: Lazaridis et al. 2025, Nature 639:132–142, Methods and Figs 1, 4, 5; Haak et al. 2015. Plain-language summary by Claude.

Example: how a percentage is built



German Corded Ware: three-way model (Haak et al. 2015, Fig. 3)

- Pick source populations (here Yamnaya, early farmers, western hunter-gatherers).
- The model finds the mix that best reproduces the target group's genetic similarities.
- Different models give slightly different numbers: 73.1% (two-way) versus 79% (three-way).
- Read each percentage with its model and its uncertainty, here ± 2.2.

Tier A: Haak et al. 2015, Nature 522:207–211, section on Corded Ware ancestry and Fig. 3.

From bone to data: the laboratory side

Clean rooms

UV, filtered air, barcoded adapters, drilling below the surface. Contamination is the main danger.

Capture, not brute force

Synthetic "bait" pulls out human DNA at over a million informative sites. Under $500 per sample.

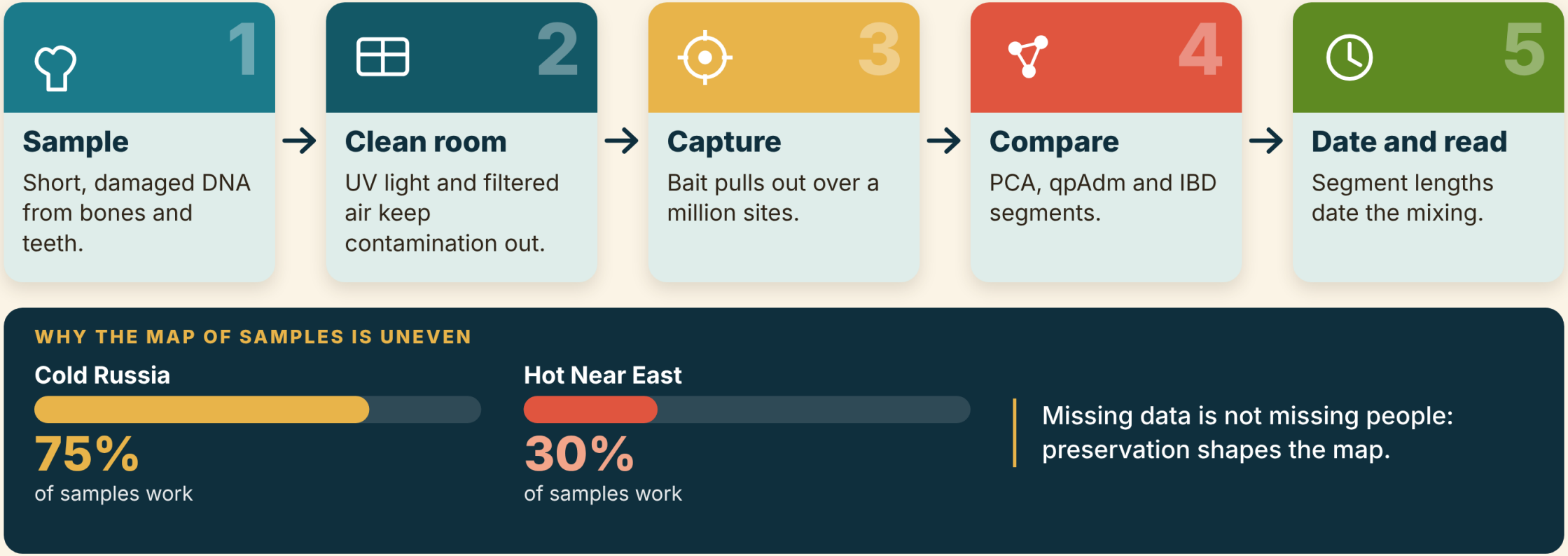
Yield depends on climate

About 75% of samples work in cold Russia, about 30% in the hot Near East.

The sample map is shaped by preservation, so absence of data is not absence of people.

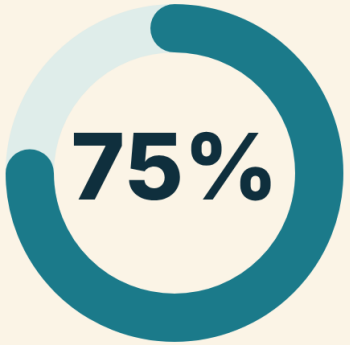
Tier A: Reich 2018, Who We Are and How We Got Here, pp. xix–xxi, 30–33 (supplied; Introduction and ch. 1–2 only). Claim that DNA beats archaeology (pp. xviii, xxii) is Reich's own position.

From bone to ancestry percentage



Tier A: Reich 2018, pp. xix–xxi and 30–33; Haak et al. 2015; Lazaridis et al. 2025, Methods.

Why the map has holes: recovery rates



Cold Russia

of screened samples give usable DNA



All regions

about half of screened skeletons



Hot Near East

of screened samples give usable DNA

< $500

per sample with bait capture

> 1 million

variable positions fished out

Why contamination is the enemy

A typical Neanderthal fragment is about **40 letters** long. Humans and Neanderthals differ at about **1 letter in 600**. A single skin flake from a handler can swamp the signal, so labs use clean rooms, UV light, filtered air, barcoded adapters and drilling below the surface.

Microbes dominate

One Neanderthal bone was only up to **4%** archaic human DNA; the rest was microbial.

Tier A: Reich 2018, pp. xix–xxi, 30–33 (digest).

Was there gene flow? Testing and dating mixture

1 Choose four groups

Two populations to compare, a third group, and an outgroup.

2 Count shared mutations

Equal counts: A and B are a clade. Unequal counts: gene flow.

3 Date it by segment length

Segments from a mixture break up each generation; long segments mean recent mixture.

Statistics say that mixing happened and roughly when. They do not say who spoke what.

Tier A: Reich 2018, pp. 34–40 (supplied). The book uses the plain name "Four Population Test"; later chapters (not supplied) develop it.

A test in action: Neanderthal ancestry

Test: who shares more mutations with a Neanderthal?



+9%

more shared mutations in Europeans; equal for East Asians and New Guineans

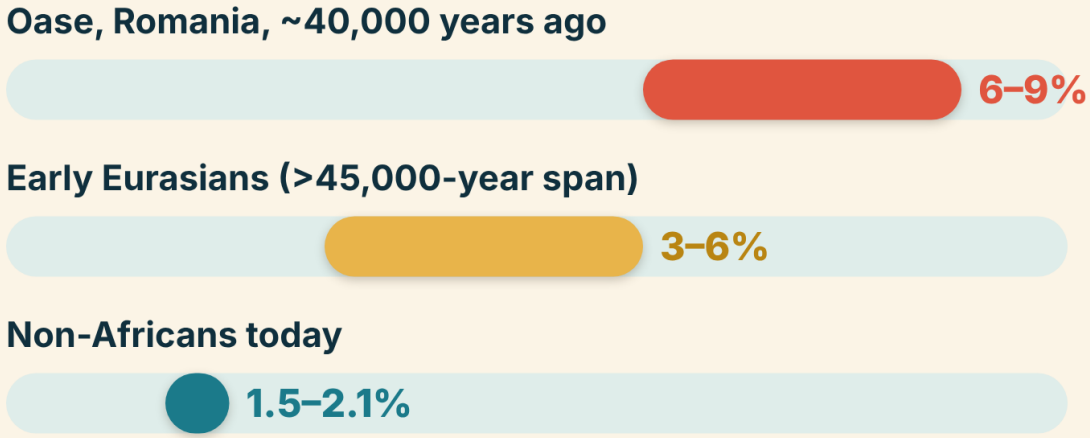
Why this counts as proof

If A and B form a clade relative to a third group, counts should be equal. Unequal counts mean **gene flow**. By chance this gap has a probability under **1 in a quadrillion**. Only positions known to vary in living humans were used, so a 0.5% sequencing error did not matter.

Dating the mixing

Segment length is a clock: a ~45,000-year-old Siberian had segments ~7× longer, narrowing entry to **54,000–49,000 years ago**.

Neanderthal ancestry, % of the genome



Tier A: Reich 2018, pp. 34–48 (digest). Percentages are ranges as stated.

Seven building blocks of Eurasian ancestry

Source population	Represented by	Time
Anatolian farmer	Western Anatolian farmers	7th millennium BCE
Iranian early farmer	Goat herders, Zagros (Ganj Dareh)	8th millennium BCE
Western hunter-gatherer	Western Europeans	9th millennium BCE
Eastern hunter-gatherer	Eastern Europeans	6th millennium BCE
West Siberian hunter-gatherer	Forest zone, central Russia	6th millennium BCE
East Siberian hunter-gatherer	Lake Baikal region	6th millennium BCE
Andamanese-related (AASI proxy)	Present-day Andaman Islanders	Present day

Tier A: Narasimhan et al. 2019, Science 365:eaat7487, Box 1. Terminology follows that paper (WEHG/EEHG = WHG/EHG elsewhere).

Seven building blocks, on the map



Reading the map

- Three hunter-gatherer groups, two farmer groups, and two more distant sources.
- Eastern and western Europe start from different hunter-gatherer ancestries.
- The steppe sits between sources 2, 3 and the farmer groups.
- Ancestry of any later group is modelled as a mixture of these blocks.

Tier A: Narasimhan et al. 2019, Science 365:eaat7487, Box 1. Base map: Natural Earth (public domain). Positions schematic.

The first farmers: Basal Eurasians and the Near East

44

ancient Near Easterners, c. 12,000–1,400 BC

~half

of earliest Near Eastern ancestry from "Basal Eurasians"

2 groups

Levant and Zagros first farmers were genetically very different

- By the Bronze Age, Levantine, Iranian and Anatolian farmers had mixed and the differences shrank.
- Iranian-related farmers spread north to the steppe and, later, east to South Asia.

Tier A: Lazaridis et al. 2016, Nature 536:419–424, abstract and opening text.

PART 2

02

Europe and the steppe

Three streams: hunter-gatherers, farmers and the Yamnaya



What ancient DNA overturned: demic diffusion

Old model

Farmers spread slowly from the Near East after about 9,000 years ago, mixing as they went (Cavalli-Sforza).

What aDNA showed

First farmers carried almost no hunter-gatherer ancestry. A later steppe migration reshaped the north.

The language step

"Probably" Indo-European languages (Reich 2018, p. xxv). A hedge, and an inference, not a measurement.

Population replacement is measured. That the incomers brought the languages is argued.

Tier A for the gradient history (Reich 2018, pp. xiv–xvii, xxi); tier C for the language inference (p. xxv). Ch. 5–6 on Yamnaya and South Asia were not in the supplied text.

Haak 2015: a migration into the heart of Europe

69

Europeans sampled, 8,000–3,000 years ago

~75%

Corded Ware ancestry traced to Yamnaya (abstract)

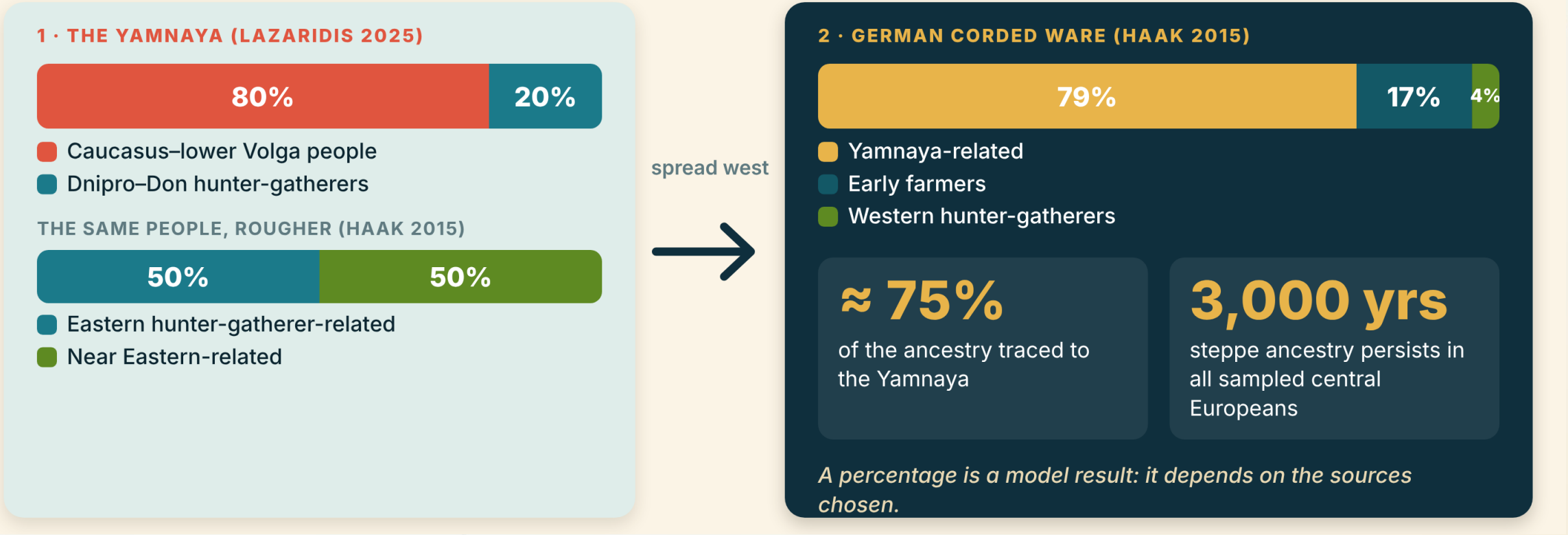
3,000 yrs

steppe ancestry persists in all sampled central Europeans

- Western and eastern Europe followed opposite trajectories between 8,000 and 5,000 years ago.
- The Yamnaya themselves were a mix: eastern hunter-gatherers plus a Near Eastern-related source.

Tier A: Haak et al. 2015, Nature 522:207–211, abstract.

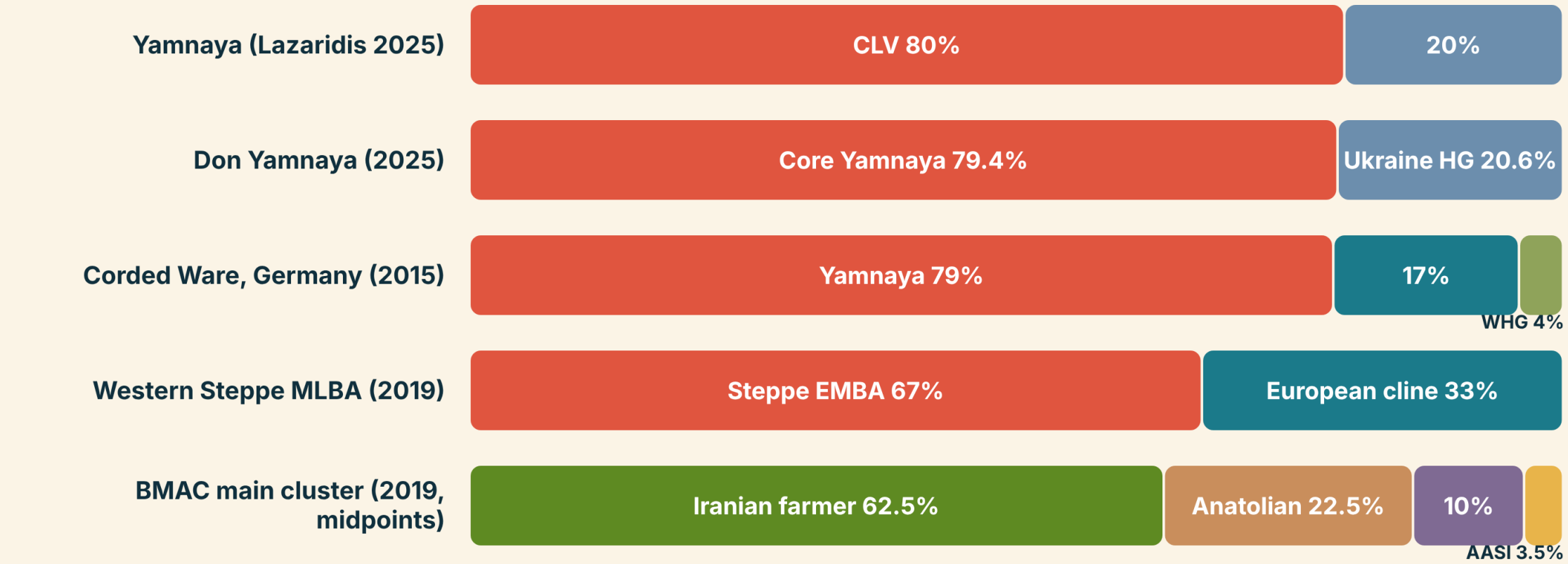
What the Yamnaya were made of, and what they became



Tier A: Haak et al. 2015 (about half); Lazaridis et al. 2025, Nature 639:132–142 (around four-fifths and one-fifth); Haak 2015, Fig. 3 (Corded Ware). Rough shares, each tied to its own reference populations.

Ancestry recipes compared

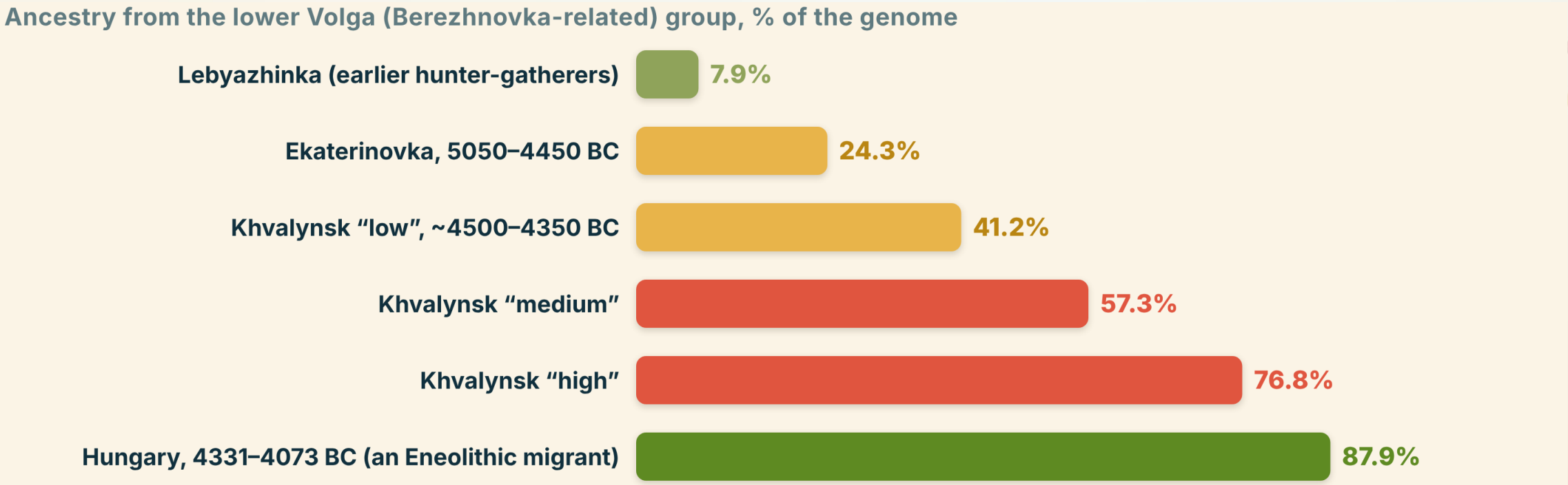
Each bar adds to 100%: the sources differ by study, so compare shapes, not exact numbers



CLV = Caucasus–lower Volga; HG = hunter-gatherer; MLBA = Middle–Late Bronze Age; AASI = Andamanese-related. BMAC values are the midpoints of the quoted ranges.

Tier A: Lazaridis 2025; Haak 2015 Fig. 3; Narasimhan 2019. BMAC bars use midpoints of the quoted ranges (Claude).

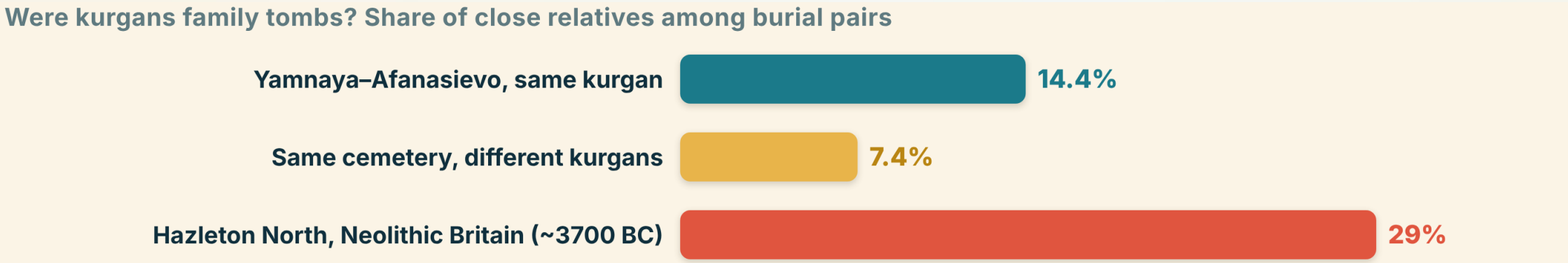
Forming the Yamnaya on the Volga



Reading: the Volga steppe was a mixing zone. Upriver hunter-gatherers (EHG) and newcomers from the lower Volga blended along a gradient of about 14–89% lower-Volga ancestry. The Yamnaya came out of this mixture, which the paper dates to about 4000 BC.

Tier A: Lazaridis et al. 2025, Nature 639:132–142 (Fig. 1c and text).

Were kurgans family tombs? Yamnaya numbers



3642–3374 BC

Yamnaya ancestors grew from a few thousand to ten times more reproducing individuals

~4000 BC

Yamnaya ancestors formed by admixture around this date

79.4 / 20.6

Don Yamnaya: Core Yamnaya vs Ukraine hunter-gatherer ancestry (%)

Reading: kurgans were not biological family tombs. Kinship inside them mostly came from common descent centuries earlier, and many close links were social, not genetic. Yamnaya sites from China to Hungary stayed genetically very similar (mean F_{ST} = 0.005).

Tier A: Lazaridis et al. 2025, Nature 639:132–142 (Fig. 4 and Extended Data).

Allentoft 2015: the Bronze Age as a time of replacements

Group	Finding
Afanasievo	Genetically indistinguishable from Yamnaya
Corded Ware and Sintashta	Close affinity; similar genetic sources
Andronovo	A temporal and geographic extension of the Sintashta gene pool
Karasuk, Mezhovskaya, Iron Age	Gradual admixture with East Asians
Okunevo	Related to present-day Native Americans (Mal'ta lineage)

Large-scale migrations and replacements shaped Bronze Age Europe and Asia.

Tier A: Allentoft et al. 2015, Nature 522:167–172, abstract and "Bronze Age Asia".

Traits: light skin early, lactase late

Light skin

Variants for light pigmentation already at high frequency in Bronze Age Europeans.

Lactase persistence

The variant for drinking milk as an adult was still rare. Selection for it came later.

Why it matters

Pastoralists did not need adult lactase to herd; a good warning against reading traits from lifestyles.

Tier A: Allentoft et al. 2015, abstract and trait analyses. Third card: commentary by Claude (tier C).

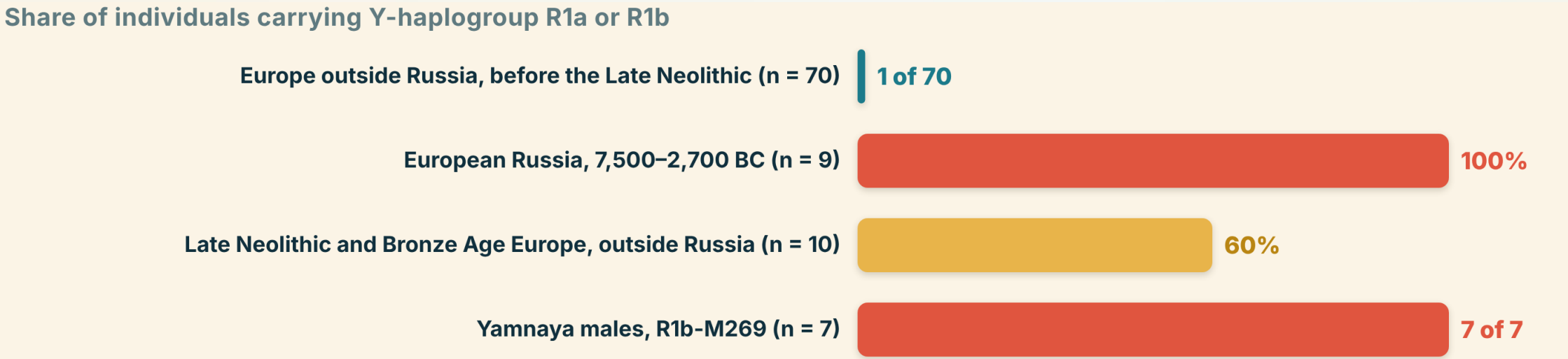
Fathers and mothers: Y-chromosomes and mtDNA

Group	Dominant Y line	Source
Early European farmers	G2a	Haak 2015
Core Yamnaya	R-M269 (49 of 51); R-Z2103 (41 of 51)	Lazaridis 2025
Don Yamnaya	I-L699 (17 of 20)	Lazaridis 2025
Bell Beaker	R-L51 prevalent	Lazaridis 2025
Late Neolithic and Bronze Age Europe	R1a and R1b replace G2a	Haak 2015

Y-lines trace fathers only. A line can spread while many other ancestries stay.

Tier A: Haak et al. 2015 and Lazaridis et al. 2025, Nature 639:132–142, Fig. 3 and Y-chromosome section.

R1a and R1b arrive from the east



R1a + R1b

most common Y-lines in many European populations today

G2a

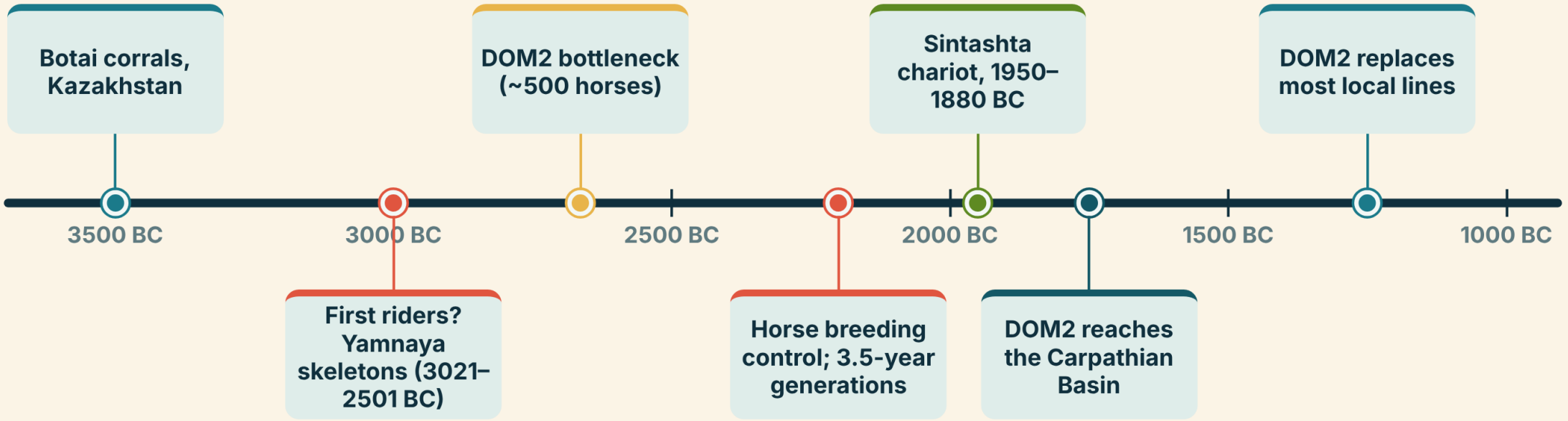
early-farmer Y-line, almost gone after the Late Neolithic

both sexes

mtDNA also shifts, so women moved too

Tier A: Haak et al. 2015, Nature 522:207–211. Small samples: read as pattern, not exact frequency.

Horses in the DNA record: a timeline



Reading: wagons came first, pulled by cattle. Large-scale horse mobility is not visible in horse DNA until about 2200 BC, so the Yamnaya push into Europe around 3000 BC was not driven by horses: Corded Ware horses lack the modern-domestic (DOM2) signature.

Tier A: Librado et al. 2021 and 2024 (Nature); Gaunitz et al. 2018; Trautmann et al. 2023; Lindner 2020.

Horse genomes: the numbers

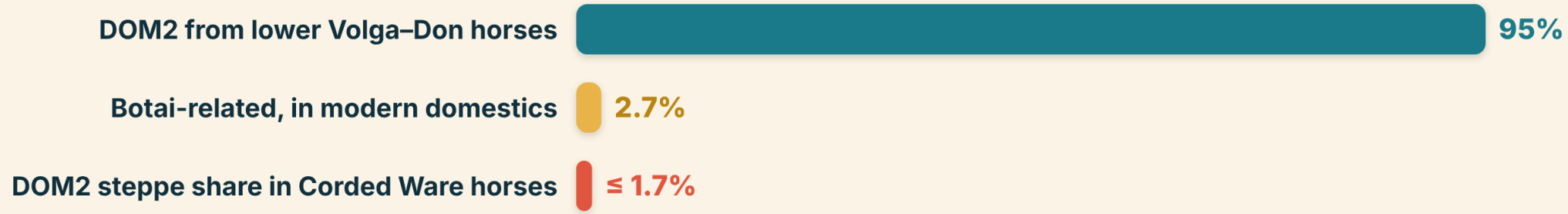
273 → 475
ancient horse genomes in the 2021 and 2024 studies

7.4 → 3.5
years per generation: a 2.1× speed-up by 2200 BC

~500
horses at the domestication bottleneck, from 2664 BC

13.7×
population expansion within 30 generations

How much ancestry did each group of horses pass to modern domestics?



Reading: the Botai horses of Kazakhstan, among the first managed in the world, left almost nothing to modern breeds, and Przewalski's horses turned out to be their feral descendants. Modern domestic horses come from one lineage of the western steppe.

Tier A: Librado et al. 2021 and 2024 (CC BY 4.0); Gaunitz et al. 2018.

Who was riding? Evidence from skeletons

Looking for riders in the skeletons (Trautmann et al. 2023)

217 individuals from 39 sites studied

150 of them Yamnaya

14 with three or more of six riding-related traits

5 meet the full threshold: 4+ traits, 7+ of 12 points

The five oldest riders

Dated 3021–2501 cal BC, from Romania, Bulgaria and Hungary. The Strejnicu man (30–40, about 165 cm) shows all six traits.

Caution

The traits are not unique to riding, and the study had no control group. Genetics says mass horse mobility came later, near 2200 BC: a tension, not a settled fact.

Tier A: Trautmann et al. 2023, Science Advances 9:eade2451.

PART 3

03

The Southern Arc



Anatolia, the Balkans and Armenia, from Lazaridis et al. 2022

Lazaridis 2022: the Southern Arc

727

ancient individuals sampled, Southern Arc, c. 10,000 years

0

steppe ancestry in Anatolia down to the Bronze Age (2022)

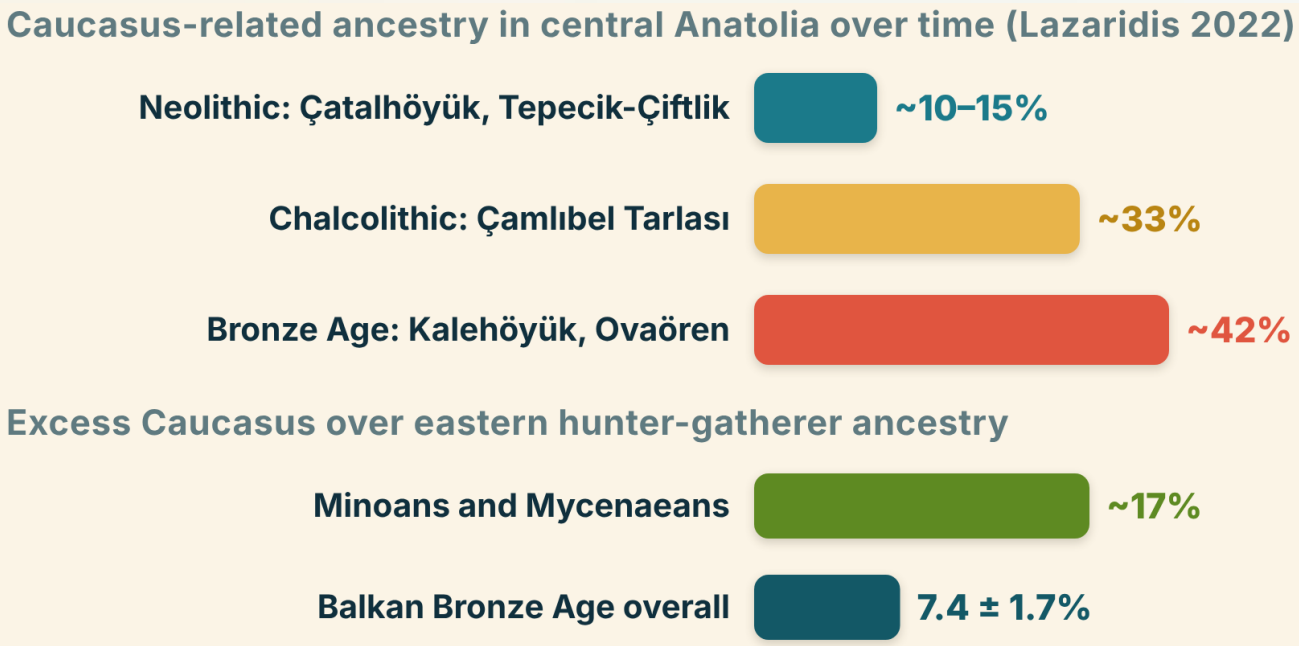
~½

of Yamnaya ancestry was West Asian

The 2022 paper placed the link between Yamnaya and Anatolian in the West Asian highlands.

Tier A: Lazaridis et al. 2022, Science 377:eabm4247, summary and abstract. "Steppe ancestry in Anatolia" is the 2022 conclusion; compare Lazaridis 2025 (CLV, about 10%).

Anatolia: a blend that became uniform



Homogenization

Differences between regions narrowed. The range of Caucasus-related ancestry shrank from about 50 to about 15 points.

No steppe ancestry in Bronze Age Anatolia, even though Indo-European languages are attested there.

Reading: the Anatolian gene pool was already a blend of local hunter-gatherers and eastern farmers well before the Bronze Age, and became more uniform. The 2022 paper placed the link between Yamnaya and Anatolian speakers in the highlands of West Asia; half of Yamnaya ancestry was West Asian.

Tier A: Lazaridis et al. 2022, Science 377:eabm4247.

The Balkans and Armenia

Balkans, 5000–4500 years ago

Some individuals traced almost all their genes to Yamnaya; later mixing with locals.

Greek, Paleo-Balkan, Albanian

Proposed to arise from the diverse mixed population that followed.

Armenia, c. 4000 years ago

A low but pervasive steppe signal; Yamnaya Y-lines persisted there.

Tier C for the language statements: the authors link gene flow to Greek and Armenian.

Tier A: Lazaridis et al. 2022, Science 377:eabm4247 summary; language links = authors' interpretation (tier C).

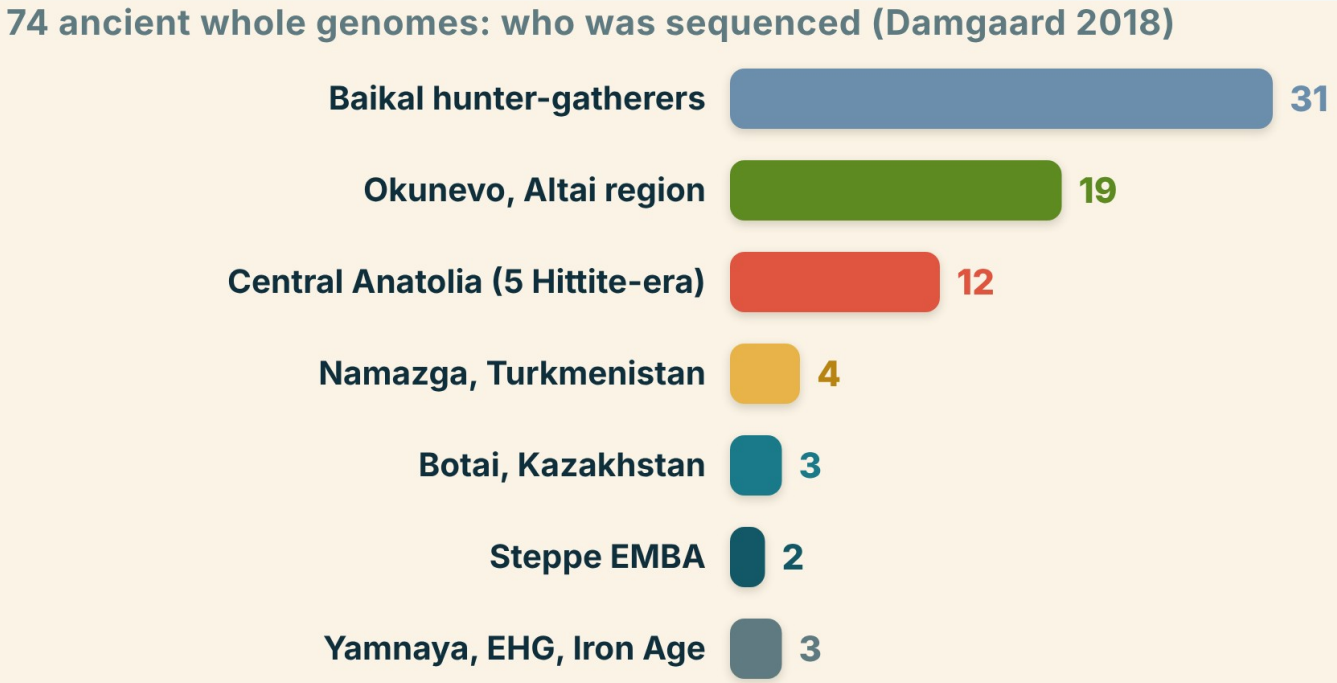
Anatolia: two readings of the same silence

Paper	Anatolia finding
Damgaard 2018	No steppe ancestry in Bronze Age Anatolia
Lazaridis 2022	No steppe ancestry; the link lay in West Asian highlands
Lazaridis 2025	CLV ancestry (at least one-tenth) in central Anatolia; proposed southward move c. 4400 BC

Not a contradiction: "steppe (Yamnaya) ancestry" and "CLV ancestry" are different references, and CLV predates the Yamnaya.

Tier A: Damgaard 2018; Lazaridis et al. 2022, Science 377:eabm4247; Lazaridis et al. 2025, Nature 639:132–142. Interpretation of the difference by Claude (tier C).

74 genomes from Inner Asia and Anatolia



Three findings

① Botai horse people came from a hunter-gatherer line deeply diverged from Yamnaya.
② West Eurasian ancestry reached South Asia before and after, not at, Yamnaya times.
③ No steppe ancestry in Anatolia while Hittite is attested.

Also in the study

41 high-coverage present-day Central Asian genomes at 30× depth, used as a modern reference for the ancient samples.

Tier A: Damgaard et al. 2018, Science 360:eaar7711 (sample list in main text).

The Caucasus: a mountain barrier that leaked

Mesolithic	Eneolithic	Early–Middle Bronze Age	Late Bronze Age
North: Eastern hunter-gatherers. South: Caucasus hunter-gatherers, with growing East Anatolian farmer ancestry.	Steppe ancestry forms; the Maykop complex links mountains and steppe. Steppe Maykop adds Siberian-related ancestry.	Pastoralist peak and territorial expansion, with long-term genetic stability.	New gene flow from several sources; steppe ancestry is absorbed into highland populations.

131

individuals from 38 sites over 6,000 years (Ghalichi 2024)

45

individuals in a 3,000-year North Caucasus transect (Wang 2019)

1,100 km

Caucasus range between the Black Sea and the Caspian: a barrier that leaked

Tier A: Ghalichi et al. 2024, Nature; Wang et al. 2019, Nat. Commun. 10:590.

PART 4

04

Central and South Asia

Narasimhan et al. 2019: 523 ancient individuals



Narasimhan 2019: the shape of the data

523

ancient individuals, 8,000 years

0%

steppe ancestry in the main BMAC cluster

up to 30%

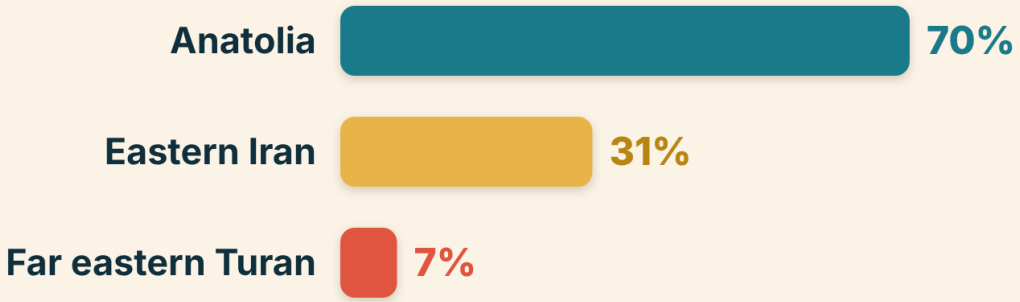
steppe ancestry in some modern South Asian groups

The steppe signal in South Asia is real, late, and entered by way of outliers and the Swat Valley.

Tier A: Narasimhan et al. 2019, Science 365:eaat7487, summary.

Anatolian-farmer ancestry fades eastward

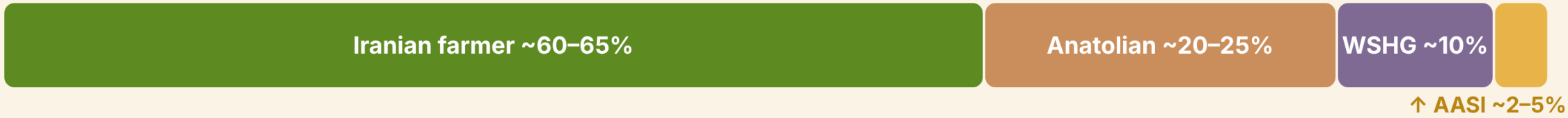
Anatolian-farmer ancestry (Narasimhan 2019)



West-to-east fade

The share of farmer ancestry from Anatolia falls from about 70% to about 7% across the Copper and Bronze Age world, a gradient that points to how far the early farmers' reach extended.

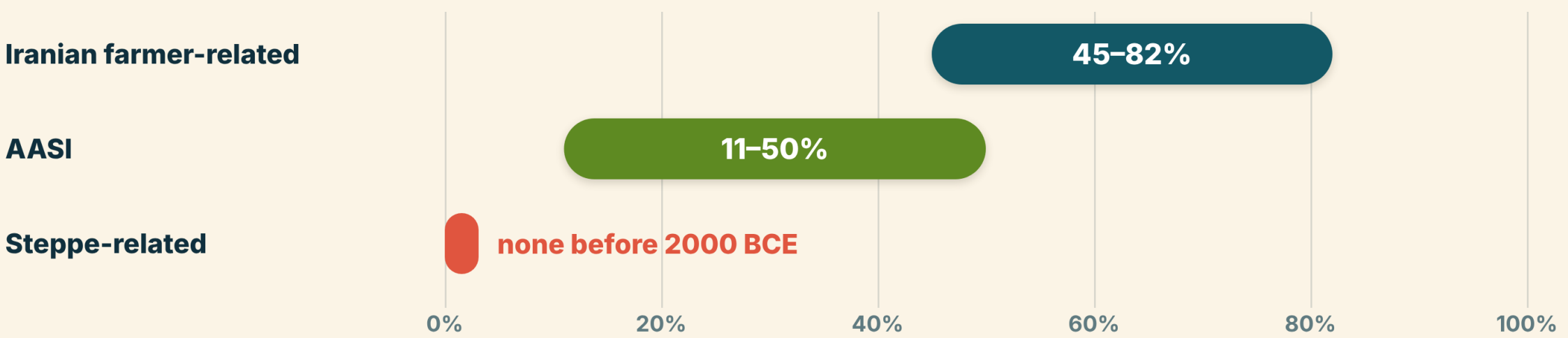
BMAC main cluster: three main sources (and a hint of the south)



The extra 2–5% Andamanese-related ancestry points to **south-to-north gene flow** from South Asia into the BMAC, in line with trade between the Indus cities and Bactria. The main BMAC cluster carries **no steppe ancestry**.

Tier A: Narasimhan et al. 2019, Science 365:eaat7487.

The Indus Periphery cline



ASI
Ancestral South Indians: at least 55% Indus Periphery ancestry.

ANI
Ancestral North Indians: at least 70% Indus Periphery, plus steppe-ancestry groups.

up to 30%
steppe ancestry in some modern South Asian groups.

Tier A: Narasimhan et al. 2019, Science 365:eaat7487. Interpretation of priestly groups = authors (tier C).

ANI and ASI: the two ends of South Asia

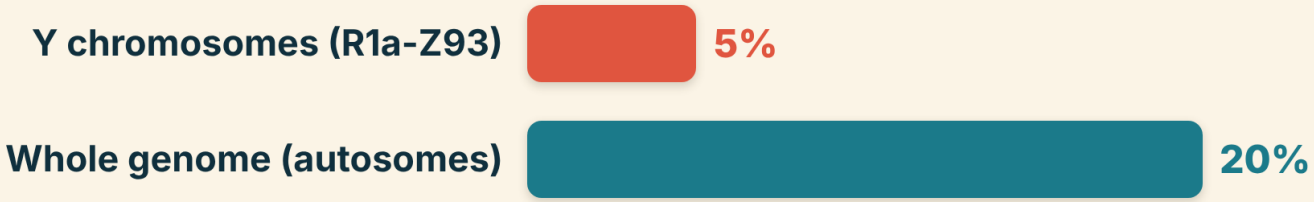
Group	Main sources (Narasimhan 2019)
ASI (Ancestral South Indians)	At least about 55% Indus Periphery; plus south-eastern groups
ANI (Ancestral North Indians)	At least about 70% Indus Periphery; plus north-western steppe-ancestry groups
Steppe ancestry	Mostly male-derived; higher in some priestly groups

Tier C: the authors read the priestly-group pattern as support for a steppe origin of Indo-Aryan.

Tier A: Narasimhan et al. 2019, Science 365:eaat7487, "South Asia" points 1–3. The interpretation is the authors' (tier C).

Sex bias: fathers and mothers moved differently

Steppe-related ancestry in Swat Valley groups, Late Bronze and Iron Age



A deficiency of males

5% of 44 Y chromosomes vs about 20% on the autosomes (Z = −3.9).

The mirror image in Europe

~90%

of early Single Grave burials in Jutland were male

28–42%

of adult women in southern German Corded Ware cemeteries were non-local

2 of 51

core Yamnaya males were not R1b-M269 (49 of 51 were)

Reading: the same steppe ancestry moved with different social stories. In South Asia it entered groups largely through **women**, in Europe the earliest migrants were largely **men**. Sex bias varied from region to region, so no single rule fits.

Tier A: Narasimhan et al. 2019; Kristiansen et al. 2017, Antiquity 91:334–347; Haak 2015 for Y-lines.

Where and when steppe ancestry shows up

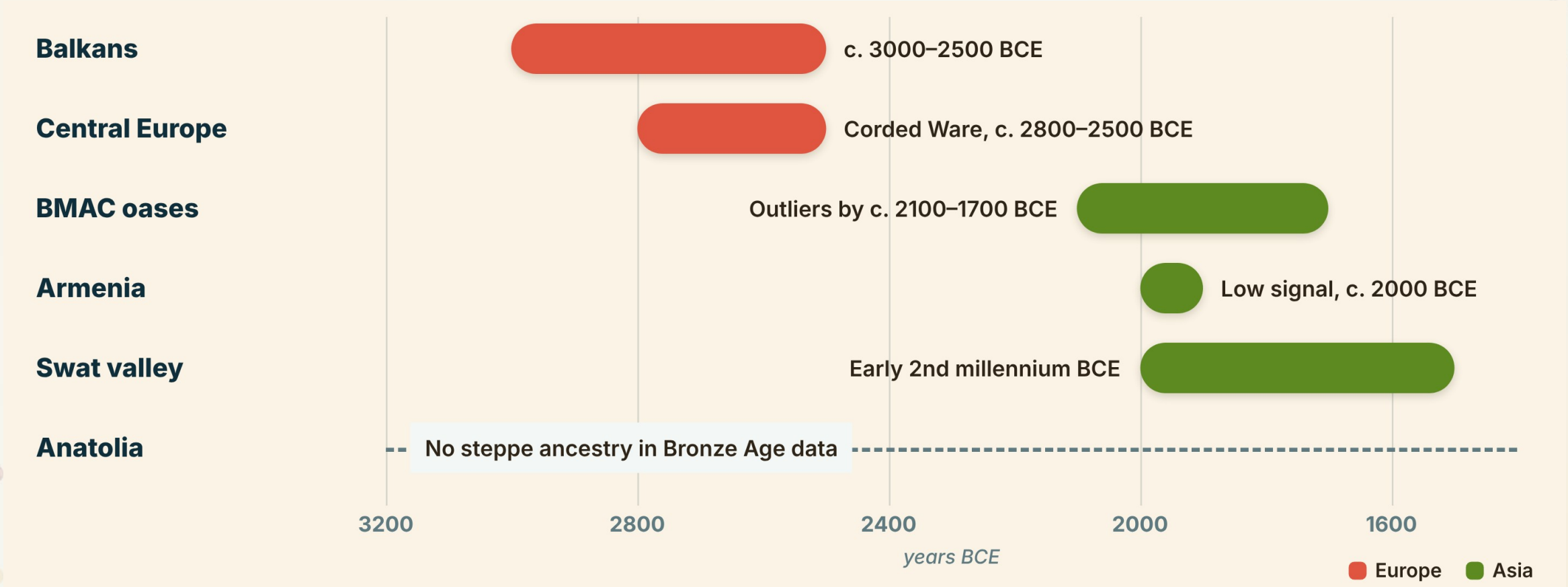


Reading the map

- Europe first: central Europe by c. 2800–2500 BCE.
- Then Asia: the BMAC oases and Swat in the early second millennium BCE.
- Anatolia is the exception: no steppe signal in the 2018 and 2022 data.

Tier A for dates (Haak 2015; Lazaridis 2022; Narasimhan 2019); positions approximate. Map: Natural Earth data.

When steppe-related ancestry appears



Tier A: Allentoft 2015; Haak 2015 (Central Europe); Lazaridis 2022 (Balkans, Armenia); Narasimhan 2019 (BMAC, Swat); Damgaard 2018 and Lazaridis 2022 (Anatolia). BCE conversion from "years ago" by Claude (about 2000 added).

PART 5

05

Limits and debates

What the genes cannot say



What DNA cannot do

Ancestry ≠ language

A people can change language without changing genes, and the reverse.

Small shares

Is 10% enough to change a language? The papers disagree.

Sparse samples

Missing evidence is not evidence of absence.

Dating errors

Freshwater fish diets can shift dates by up to a millennium.

Tier A for the dating caution (Lazaridis et al. 2025, Nature 639:132–142); other points tier C (general method caution).

Three misreadings to avoid

THE MISREADING

"DNA shows where the Indo-European homeland was."

THE BETTER READING

DNA shows where people lived and moved. It cannot show what language they spoke; that gap is why the debate continues.

THE MISREADING

"About 10% steppe ancestry means the language was replaced."

THE BETTER READING

The papers disagree. A people can change language without changing genes, and the reverse; small shares need other evidence.

THE MISREADING

"No DNA from a region means nobody lived there."

THE BETTER READING

The sample map follows preservation. About 75% of samples work in cold Russia, about 30% in the hot Near East.

Tier A/C: Reich 2018; Lazaridis et al. 2025; Heggarty 2025 (commentary). Synthesis by Claude.

Five questions for any DNA headline

1. **Which study?**
   Check the paper, the year and how many individuals were sampled.

2. **Which sources?**
   A percentage is tied to the reference populations chosen for the model.

3. **Which uncertainty?**
   Look for the ± and for how the date was estimated.

4. **Which place and time?**
   A sample is a few people from one site, not a whole region.

5. **Ancestry or language?**
   Ancestry is measured; what people spoke is inferred.

Synthesis by Claude from the supplied papers.

Two readings of Lazaridis 2025

Question	Lazaridis et al. 2025	Heggarty 2025 (commentary)
Where did Indo-Anatolian unite?	In CLV people, north of the Caucasus, 4400–4000 BC	In the West Asian highlands (Caucasus/Zagros); the key movement runs northward
What is "Indo-European"?	Non-Anatolian branches; whole family = Indo-Anatolian	Objects to renaming; Anatolian belongs to the family
Do other branches come from the steppe?	Yes, via Yamnaya and descendants	Case rests on low steppe percentages in several regions
Language evidence	Cited from linguists	"No linguistic data" in the paper

Tier C: positions as stated in Lazaridis et al. 2025, Nature 639:132–142 and Heggarty 2025 (online commentary). Both views to be read in full.

What most sources agree on

Agreed (tier A)

Yamnaya spread, Corded Ware turnover, Afanasievo, Sintashta–Corded Ware kinship, BMAC without steppe ancestry.

Disputed (tier C)

Where Indo-Anatolian began, and what steppe percentages mean for language.

Next: language

Module 4 asks what the languages themselves say about the family tree and its dates.

Synthesis by Claude from the supplied papers.

Key takeaways

- **Methods.** Ancestry percentages come from models; always ask which sources and what uncertainty.
- **Europe.** A migration from the steppe changed central Europe c. 2800–2500 BCE; Corded Ware carries about three-quarters Yamnaya-related ancestry.
- **The Yamnaya.** About four-fifths CLV, about one-fifth Dnipro–Don hunter-gatherers (Lazaridis 2025).
- **The south.** Anatolia shows little or no Yamnaya ancestry; the BMAC main population none; South Asia gets steppe ancestry late and mostly through males.
- **Language.** DNA moves; languages are inferred. The origin of Indo-Anatolian is still debated.

Summary of Module 3.

Check your understanding

1. What is “capture” in an ancient-DNA study, and how many sites does it target?
2. About what share of German Corded Ware ancestry traced to the Yamnaya in Haak 2015?
3. What two components made up the Yamnaya in Lazaridis 2025?
4. What did Narasimhan 2019 find in the main BMAC cluster?
5. Why are samples from the hot Near East scarcer than from cold Russia?
6. True or false: ancestry percentages tell us which language a group spoke.
7. Which Y-chromosome line dominated the core Yamnaya?
8. How do Damgaard 2018 and Lazaridis 2025 differ on Anatolia?
9. In Swat, what share of Y chromosomes versus autosomes was steppe-related?
10. When does horse DNA show reproductive control of the modern lineage, and what happened to generation time?

Questions drawn from the whole module.

Check your understanding: answers

1 **Capture**
Synthetic bait enriches human DNA at about 400,000 to 1.2 million known variable sites.

2 **About 75%**
Abstract of Haak 2015; 79% Yamnaya-related in the three-way model.

3 **CLV and Dnipro–Don**
Caucasus–lower Volga people (about 80%) and Dnipro–Don hunter-gatherers (about 20%).

4 **Zero steppe ancestry**
The main BMAC cluster had none; steppe signal arrives later by outliers and Swat.

5 **Preservation**
About 75% of samples yield DNA in cold Russia, about 30% in the hot Near East.

6 **False**
Ancestry is measured; language is inferred and can change independently.

7 **R-M269**
49 of 51 core Yamnaya males; R-Z2103 in 41 of 51.

8 **Reference changed**
Damgaard found no steppe ancestry; Lazaridis 2025 finds at least one-tenth CLV ancestry in central Anatolia.

9 **5% vs about 20%**
Y chromosomes of steppe type (R1a-Z93) were far rarer than steppe ancestry on the autosomes: a female route.

10 **About 2200 BC**
Generation time fell from about 7.4 to 3.5 years in the DOM2 lineage.

Tier A: Haak 2015; Lazaridis 2025; Narasimhan 2019; Damgaard 2018; Reich 2018.

Key terms

Ancient DNA
DNA recovered from the remains of people who died long ago; short, damaged and mostly microbial.

SNP
A single DNA position that varies between people. Panels read 400,000 to 1.2 million of them.

Capture
Enriching known sites with synthetic “bait” before sequencing.

PCA
A map of genetic similarity: mixed groups fall between their sources.

qpAdm
Tests which mixture of source populations fits a target group; gives percentages.

IBD segment
A long stretch of DNA shared through a recent common ancestor.

Admixture
Mixing of populations that had been separate; dated from segment lengths.

Cline
A gradient of ancestry between two ends, across space or among groups.

Y-line, mtDNA
Y-chromosome lineages trace fathers only; mitochondrial DNA traces mothers only.

CLV
Caucasus–lower Volga cline: the Caucasus hunter-gatherer-rich source of the Yamnaya.

WHG, EHG
Western and eastern hunter-gatherers: two early European ancestry sources.

AASI
Ancestry related to South Asian hunter-gatherers; Andamanese Islanders are a proxy.

Definitions summarised from the supplied papers and Reich 2018.

More key terms

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

Corded Ware
Late Neolithic culture of northern and central Europe, with about 75% Yamnaya-related ancestry.

Kurgan
A burial mound, typical of the steppe from the Eneolithic on.

BMAC
Bactria–Margiana Archaeological Complex: oasis towns in Central Asia, about 2300–1600 BC.

Indus Periphery
Outlier individuals near the Indus cities with Iranian-farmer and AASI ancestry.

DOM2
The horse lineage that gave all modern domestic horses.

Maykop
Caucasus Bronze Age complex that linked the mountains with the steppe.

Sex-biased admixture
Gene flow where one sex contributes more; read from Y chromosomes vs autosomes.

Homogenization
Regions become genetically more alike over time, as in Bronze Age Anatolia.

Reference population
A group chosen as a source in a model; changing it changes the percentages.

Eneolithic
Copper Age: roughly 5000 to 3500 BC on the steppe.

Haplogroup
A branch of the Y-chromosome or mtDNA tree, such as R1b-M269.

Definitions summarised from the module slides and the papers.

Sources for Module 3

- Allentoft, M. E. 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.
- Haak, W. et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211.
- Heggarty, P. 2025. Commentary on Lazaridis et al. 2025 (online).
- Lazaridis, I. et al. 2016. Genomic insights into the origin of farming in the ancient Near East. Nature 536:419–424.
- Lazaridis, I. et al. 2022. The genetic history of the Southern Arc. Science 377:eabm4247.
- Lazaridis, I. et al. 2025. The genetic origin of the Indo-Europeans. Nature 639:132–142.
- Narasimhan, V. M. et al. 2019. The formation of human populations in South and Central Asia. Science 365:eaat7487.
- Reich, D. 2018. Who We Are and How We Got Here. Pantheon (Introduction and ch. 1–2 read; ch. 3–12 not yet).
- Map: Natural Earth (public domain). Tier B/C items are marked on each slide.

To add when readable copies arrive: Reich 2018 ch. 5–6; Olalde 2018; Skoglund; Kristiansen & Larsson 2005.

More sources for Module 3

- Librado, P. et al. 2021. The origins and spread of domestic horses from the Western Eurasian steppes. Nature 598:634–640 (CC BY 4.0).
- Librado, P. et al. 2024. Widespread horse-based mobility arose around 2200 BCE in Eurasia. Nature 631:819–825 (CC BY 4.0).
- Gaunitz, C. et al. 2018. Ancient genomes revisit the ancestry of domestic and Przewalski’s horses. Science 360:111–114.
- Trautmann, M. et al. 2023. First bioanthropological evidence for Yamnaya horsemanship. Science Advances 9:eade2451 (CC BY).
- Lindner, S. 2020. Chariots in the Eurasian steppe: a Bayesian approach to the emergence of horse-drawn transport. Antiquity 94:361–380.
- Kristiansen, K. et al. 2017. Re-theorising mobility and the formation of culture and language among the Corded Ware culture in Europe. Antiquity 91:334–347.
- 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.
- 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 (open access).