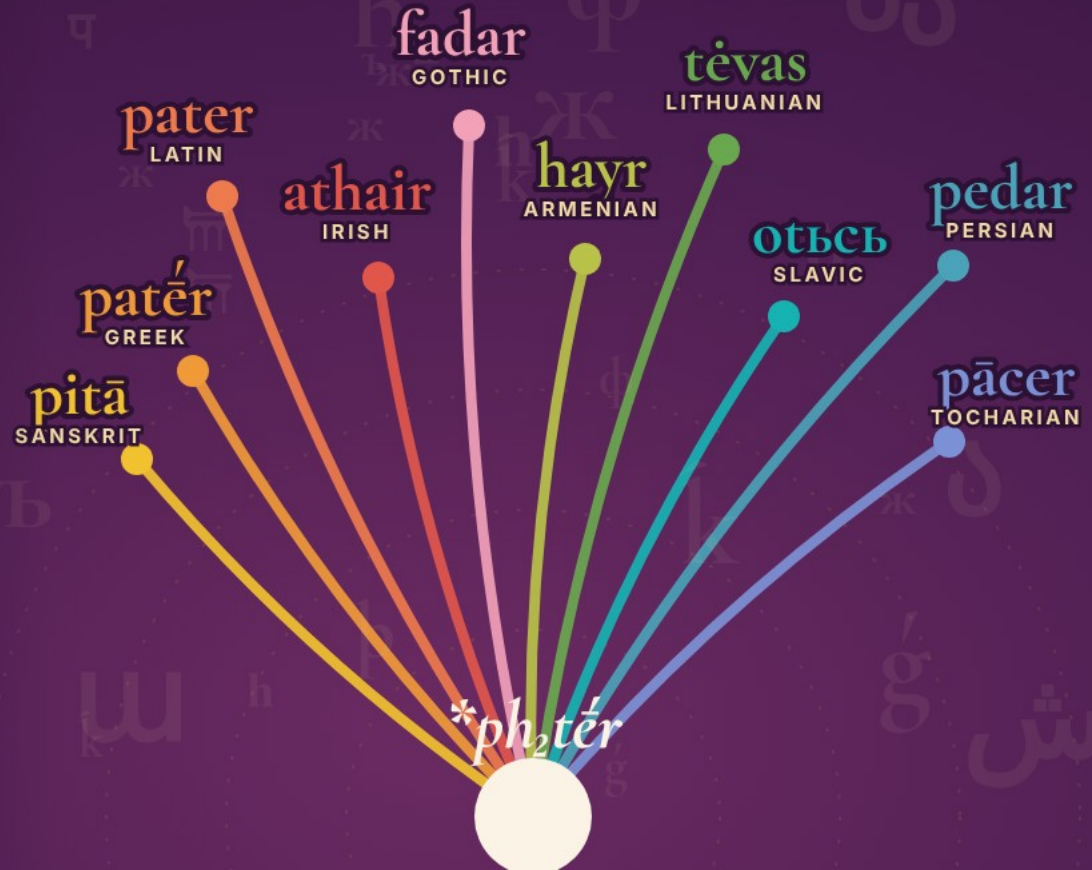
INDO-EUROPEAN COURSE · MODULE 4

04

How We Know a Family: The Comparative Method

Cognates, sound laws, reconstruction, the ten branches, and the dispute over dates



YSATIS

@airyanematlas · airyanematlas.github.io

What this module covers

1. **The method**
   Cognates, regular sound laws, reconstruction

2. **Famous laws**
   Grimm, Verner, aspirates, labiovelars, laryngeals

3. **Words that match**
   Mother, numbers, "is", fire, water, wheel

4. **The branches**
   Ten main groups and their oldest records

5. **Trees and dates**
   How dates are estimated, and the dispute

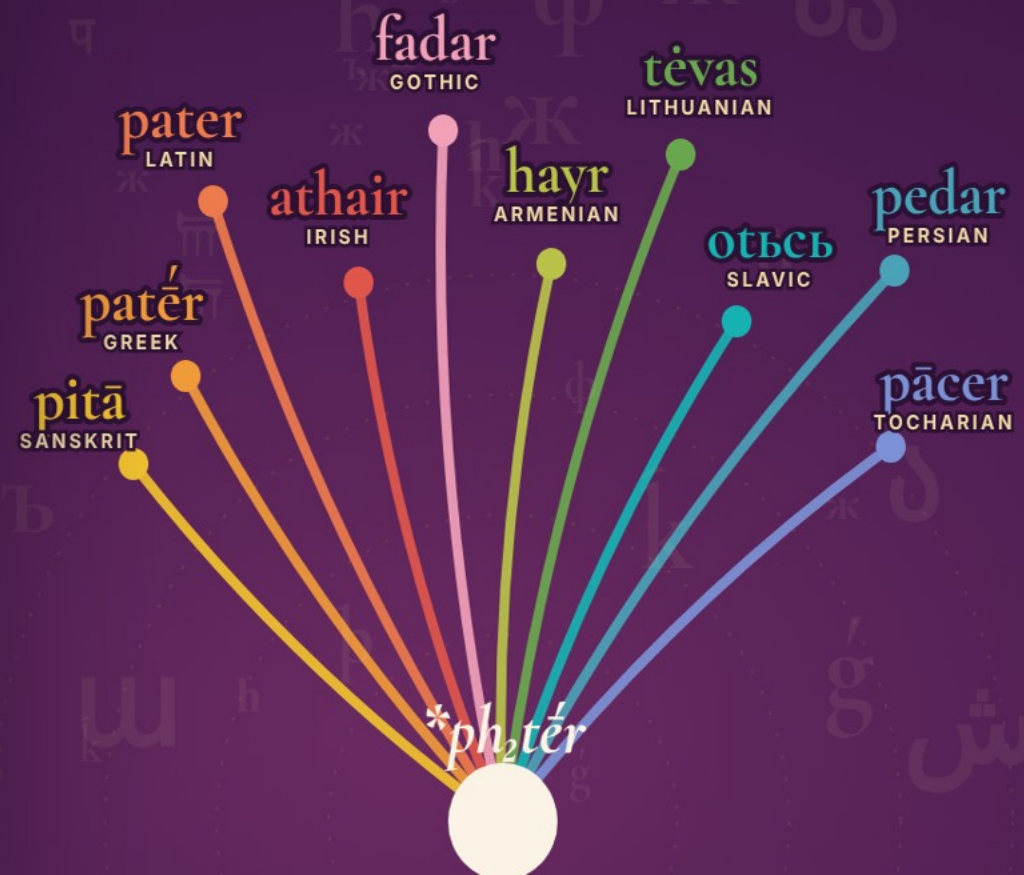
6. **Language, genes, contact**
   What each kind of evidence can and cannot show

PART 1

01

The method

Cognates, sound laws and famous exceptions



The 1786 sentence

> *"...a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident."*
>
> Sir William Jones, 1786, on Sanskrit, Greek and Latin (public domain)

- Jones saw that the likeness lay in verb roots and grammar, not in a few look-alike words.
- The next 200 years turned that insight into a method that can reconstruct unwritten languages.

Source: Jones, Asiatick Researches vol. 1 (1788; discourse of 1786), public domain. Quotation to verify against a scan (Drive library item S01).

What a language family is

Shared ancestor

Related languages descend from an earlier language that nobody wrote down: Proto-Indo-European (PIE).

Regular match

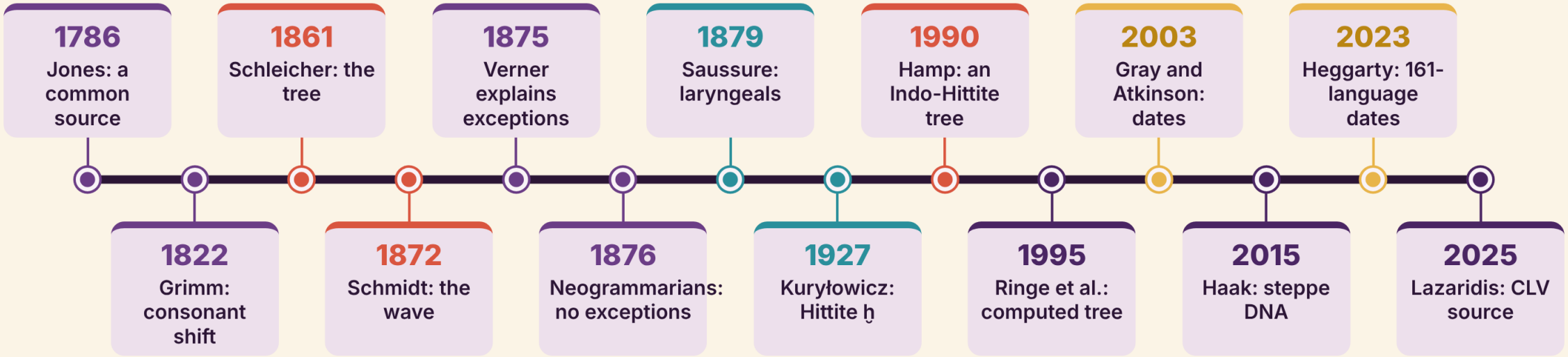
The proof is regular correspondences of sound, such as Latin p with English f in pater / father, pes / foot.

Reconstructed

PIE is rebuilt by comparison and marked with an asterisk: *ph₂tḗr "father".

The family spans Europe, Iran and South Asia, with extinct branches in Anatolia and Central Asia.

240 years of Indo-European linguistics



Three kinds of progress. Models of how languages split (tree, wave, computational). Sound laws that turned resemblance into proof. Predictions that came true, such as the laryngeals confirmed by Hittite 48 years later. Since 2015 the field has also had DNA to compare with its trees.

Tier A/B: dates of Grimm, Verner, Leskien, Saussure and Kuryłowicz from slide 8; trees from Mallory and Adams 2006 ch. 5; 1872 Schmidt, 2003 Gray and Atkinson: standard; 2015 Haak, 2023 Heggarty, 2025 Lazaridis as cited in this deck.

The comparative method in five steps

1

Collect

Match words for basic meanings across languages.

2

Correspond

Find regular sound correspondences, not stray resemblances.

3

Reconstruct

Work back to the proto-sound and proto-word.

4

Test

Check against typology, other words and attested old forms.

5

Group

Subgroup by shared innovations, not by shared retentions.

The key idea: only sound correspondences that recur across many words count as evidence.

Standard method (tier A); see Fortson; Clackson; Beekes (Drive library K01, K07, K08).

Cognate, loan or chance?

Cognate

Descends from a common ancestor. Recurs in many words with regular sound shifts: pedar ~ father; pā ~ foot.

Loan

Copied from another language. Often a cultural word and irregular in sound: English paradise from Old Iranian *pari-daiza-.

Chance

Looks alike by accident and does not recur: Persian bad and English bad are unrelated.

Mama, papa and other baby-talk words are weak evidence: they arise independently.

Standard method (tier A). Persian bad / English bad: to verify in Wiktionary and Hassandoust (Drive library).

Regular sound laws

1822	**Grimm**	Germanic consonant shift
1875	**Verner**	Accent explains the exceptions
1876	**Leskien and the Neogrammarians**	Sound laws have no exceptions
1879	**Saussure**	Laryngeals predicted
1927	**Kuryłowicz**	Hittite ḫ matches the laryngeals

Standard history of the method (tier A): Fortson; Clackson.

Grimm's Law at a glance

*p	f	Latin → English	pēs → foot
*t	th	Latin → English	trēs → three
*ḱ	h	Latin → English	cornū → horn
*d	t	Latin → English	duo → two
*g	k	Latin → English	ager → acre
*b^h	b	Latin → English	frāter → brother

Tier A, standard. Examples to verify in Fortson and Ringe; Persian parallels in Module 5.

Verner's Law: the exception explained

Sanskrit pitá́ → Gothic fadar

accent after the t → voiced d

Sanskrit bhrā́tā → Gothic brōþar

accent before the t → voiceless þ

- Same PIE sound (*t), two outcomes in Germanic.
- The difference was the place of the accent, which Sanskrit still shows.
- A puzzle solved by comparison: this is why the method is trusted.

Tier A, standard. Examples to verify in Fortson / Ringe.

Voiced aspirates in four branches

PIE	Sanskrit	Greek	Latin	English	Example
***bʰ**	bh	ph (φ)	f	b	bhrātā / phrā́tēr / frāter / brother
***dʰ**	dh	th (θ)	f	d	dvā́r- / thúra / forēs / door
***gʰ**	gh / h	kh (χ)	h	g	haṃsá- / khḗn / (h)ānser / goose

Persian turned them into plain voiced stops (b, d, g): barâdar "brother", dar "door". See Module 5.

Tier A for brother and door sets; verify "goose" and the Latin reflex in Beekes / de Vaan / Fortson.

“Four”: one sound, three outcomes

***kʷ**

“four”

one source sound, three outcomes

kʷ kept as k / kw	kʷ became p	kʷ went its own way
Latin **quattuor**	Oscan **petora**	Sanskrit **catvā́raḥ (c)**
Old Irish **cethir**	Welsh **pedwar**	Greek **téttares (t)**
		English **four (f)**

Forms standard (tier A); treatment in English and Greek simplified (tier B). Verify in Fortson, Beekes.

Centum and satem: the “hundred” test

*ḱm̥tóm “hundred”: the palatovelar *ḱ splits the family

centum branches · *ḱ became k or h		satem branches · *ḱ became s, š or ś	
Latin	**centum**	Sanskrit	**śatám**
Greek	**(he)katón**	Avestan	**satəm**
Old Irish	**cét**	Persian	**sad**
Gothic	**hund**	Lithuanian	**šimtas**
Tocharian B	**känt**	Old Church Slavonic	**sŭto**

Reading: satem and centum is a pattern of sound change, not a family tree. The satem languages are found in the east and centre (Indo-Iranian, Balto-Slavic, Albanian, Armenian), centum in the west and at both ends (Celtic, Germanic, Italic, Greek, Tocharian, Anatolian). Finnish sata “100” was borrowed from an Indo-Iranian form.

Tier A, standard forms. Verify in Fortson and Mallory and Adams 2006 (Indo-Iranian and Balto-Slavic share satemization and the ruki rule).

Laryngeals: Saussure's prediction

1879

Saussure predicts extra sounds in PIE from patterns in vowels, long before any text showed them.

1927

Kuryłowicz reads the Hittite sound ḫ as one of these sounds when Hittite is deciphered.

Example

*$h_2$ent-: Hittite ḫanti, Greek antí, Latin ante, Gothic and-.

A prediction from pure comparison, confirmed by a newly deciphered language.

Tier A history; example forms standard, verify in Kloekhorst 2008 (Drive L09).

Reading a reconstruction

Symbol	Meaning	Example
*	reconstructed, not attested	*ph₂tḗr "father"
h₁ h₂ h₃	the three laryngeals	*h₂ent- "front"
ḱ ǵ ǵʰ	palatovelars (become s/z in satem languages)	*ḱm̥tóm "hundred"
bʰ dʰ gʰ	voiced aspirates	*bʰréh₂tēr "brother"
r̥ l̥ m̥ n̥	syllabic resonants	*ḱm̥tóm; *h₃nómn̥
e / o / ø	ablaut (vowel alternation)	*bʰer- / *bʰor- / *bʰr̥-

*Standard notation (tier A). *h₃nómn̥ "name" listed for illustration; verify in Fortson / Beekes.*

Reading a starred form: *bʰréh₂tēr

*bʰréh₂tēr

*****

a hypothesis, not a recorded word

bʰ

voiced aspirate: Sanskrit bh, Greek ph, Latin f, Germanic b

é

accent: Sanskrit shows it; Verner's Law needs it

h₂

laryngeal: colours e to a, giving Sanskrit ā and Latin ā

-tēr

kin suffix, as in *ph₂tḗr "father"

Outcomes

Sanskrit
bhrā́tā

Greek
phrā́tēr

Latin
frāter

Gothic
brōþar

Old Church Slavonic
bratrŭ

Tier A, standard. Verify in Fortson, Beekes.

How reliable is a reconstruction?

1%

of 1,474 reconstructions have cognates in all twelve major groups

75%

rest on cognates in six groups or fewer

15

words are attested in every group

- **In every group:** water, cow, foot, door, three, five, seven, nine, sleep, name, I, we, thou, ye, that.
- About half of all reconstructions rest on only four or five groups.
- A starred form is a hypothesis with a pedigree: ask how many branches support it.

Tier A: Mallory and Adams 2006, pp. 107–108, Tables 7.1–7.2 (supplied; paraphrased). Counts refer to their 1997 dictionary.

Fifteen words found in every group

Nature

- water

Animals

- cow

Body and home

- foot
- door

Numbers

- three
- five
- seven
- nine

Life

- sleep
- name

Pronouns

- I
- we
- thou
- ye
- that

15

words with cognates in every one of the twelve major groups

1%

of 1,474 reconstructions (the 1997 dictionary) reach all twelve groups

75%

rest on six groups or fewer; about half on only four or five

Tier A: Mallory and Adams 2006, ch. 7.2, pp. 107–108, Tables 7.1–7.2 (supplied; paraphrased).

Labels that keep a reconstruction honest

Where the cognates are	Label	How it is treated
Celtic, Italic, Germanic, Baltic, Slavic only	North-West (NW)	Not necessarily PIE
Albanian, Greek, Armenian, Phrygian only	Central (C)	Not necessarily PIE
Anatolian plus any other group	—	Accepted as PIE
A European group plus Indo-Iranian or Tocharian	—	Accepted as PIE
Greek plus Indo-Iranian only	GA	Flagged as a special case
Indo-Iranian plus Tocharian only	Eastern (E)	Labelled as an eastern pattern

These are one pair of authors' working rules; other scholars set the bar differently.

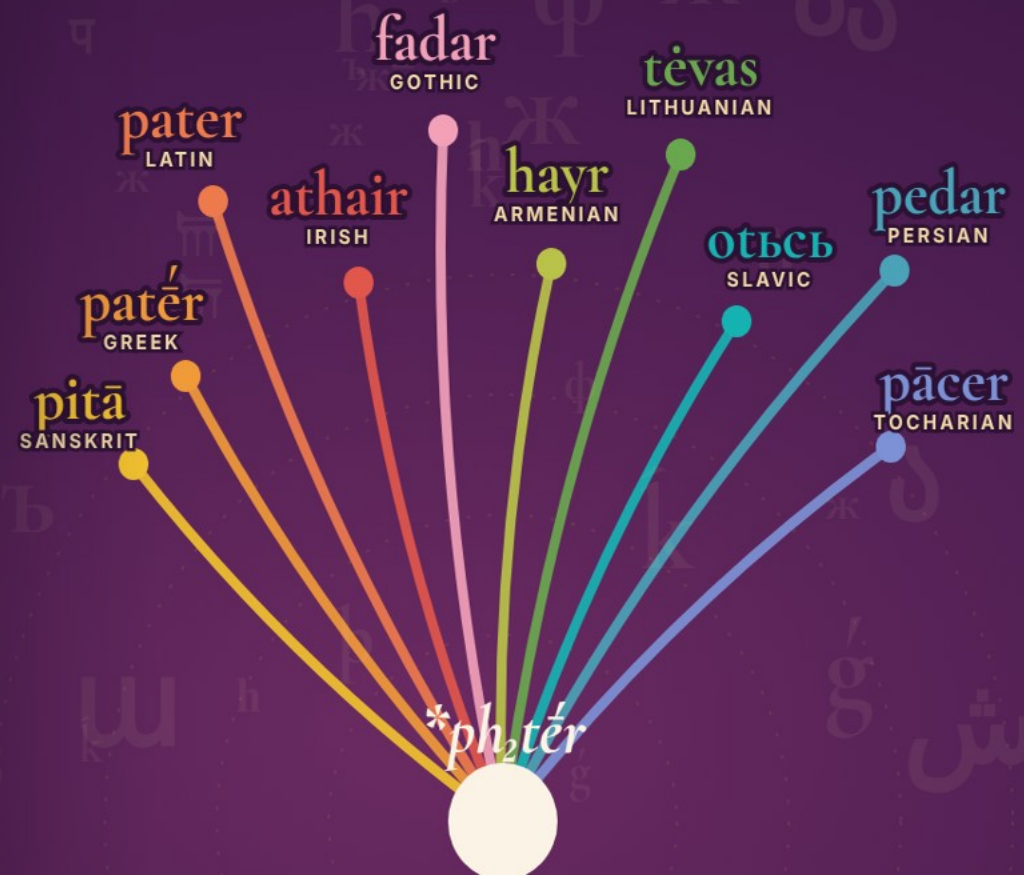
Tier A: Mallory and Adams 2006, pp. 108–110 and Fig. 7.1 (supplied; paraphrased).

PART 2

02

Words that match

Mother, numbers, “is”, fire, water and the wheel



"Mother" across eight branches

Branch	Form	Meaning
Indo-Iranian	Skt mātár-; Persian mādar	mother
Greek	mḗtēr (Doric mā́tēr)	mother
Italic	Latin māter	mother
Celtic	Old Irish máthir	mother
Germanic	Old English mōdor	mother
Slavic	Old Church Slavonic mati	mother
Armenian	mayr	mother
Tocharian	A mācar; B mātar	mother
Albanian	motër	"sister": the meaning shifted

Tier A for the set; Albanian motër "sister" standard. Verify in Fortson and Orel (Drive library K01, L17).

Family words: father, brother, sister, daughter

father *ph₂tḗr		brother *bʰréh₂tēr		sister *swésōr		daughter *dʰugh₂tḗr	
Sanskrit	**pitā́**	Sanskrit	**bhrā́tā**	Sanskrit	**svásā**	Sanskrit	**duhitā́**
Greek	**patḗr**	Greek	**phrā́tēr***	Latin	**soror**	Greek	**thugátēr**
Latin	**pater**	Latin	**frāter**	Old Irish	**siur**	Gothic	**daúhtar**
Old Irish	**athair**	Old Irish	**brāthir**	Gothic	**swistar**	Lithuanian	**duktẽ**
Gothic	**fadar**	Gothic	**brōþar**	Lithuanian	**sesuõ**	Armenian	**dustr**
Armenian	**hayr**	Armenian	**ełbayr**	Armenian	**kʻoyr**	Tocharian B	**tkācer**
Tocharian B	**pācer**	Tocharian B	**procer**	Tocharian B	**ṣer**	Persian	**doxtar**
Persian	**pedar**	Persian	**barādar**	Persian	**xwāhar**	Church Slavonic	**dъšti**

* Greek phrā́tēr means a member of a clan-brotherhood; “brother” is adelphós. Latin sister survives; daughter is replaced by fīlia.

Tier A/B, standard forms; verify in Fortson, Beekes, Mallory and Adams ch. 12 (not yet read). Persian forms: Module 5.

Numbers across ten branches

Language	two	three	five	seven	ten
Sanskrit	dvá	tráyas	páñca	saptá	dáśa
Persian	do	se	panj	haft	dah
Greek	dúo	treîs	pénte	heptá	déka
Latin	duo	trēs	quīnque	septem	decem
Old Irish	dá	trí	cóic	secht	deich
Gothic	twai	þreis	fimf	sibun	taíhun
Lithuanian	dù	trys	penki	septynì	dẽšimt
Old Church Slavonic	dъva	trije	pętь	sedmь	desętь
Armenian	erku	erek‘	hing	ewt‘n	tasn
Tocharian B	wi	trai	piś	ṣuk(t)	śak

Standard forms, tier A/B. Verify in Fortson, Beekes, Adams (Tocharian B). Persian row: Module 5 (verified).

“Seven”: a regular exception

Sanskrit	Latin	Old Irish	Gothic	Lithuanian
saptá	**septem**	**secht**	**sibun**	**septynì**
starts with s	starts with s	starts with s	starts with s	starts with s

Old Church Slavonic	Persian	Greek	Armenian	Tocharian B
sedmь	**haft**	**heptá**	**ewtʻn**	**ṣuk(t)**
starts with s	starts with h	starts with h	starts with ew-	starts with ṣ

s is kept in most branches; Persian and Greek turned it into **h**: a regular change.

Standard forms, tier A/B. Verify in Fortson, Beekes, Adams (Tocharian B).

"Is": one verb, many branches

Language	"is"
Sanskrit	ásti
Persian	ast
Greek	esti
Latin	est
Gothic	ist
Hittite	ēšzi (ēš-zi)
Lithuanian	ẽsti
Old Church Slavonic	jestь

*PIE *h₁ésti: even the most irregular verb is regular in its correspondences.*

Tier A for the set; Hittite form to verify in Kloekhorst (Drive L09).

Fire and water: an old pattern

Meaning	Hittite	Greek	Sanskrit	English	Tocharian
water	watar (gen. wedenas)	húdōr (gen. hú-d-atos)	udán- (udakám)	water	—
fire	pāḫḫur (gen. paḫḫuenaš)	pûr	agní- (a different word)	fire	A por / B puwar

Anatolian keeps archaic features that the other branches changed.

Tier B. Forms standard; verify in Fortson and Kloekhorst 2008. Sanskrit agni: see Module 5 (Rigveda 1.1.1).

Wheel words and the date of the family

Language	"wheel" or "wagon"
Sanskrit	cakrá-
Greek	kúklos
English	wheel
Tocharian A / B	kukäl / kokale "wagon"
Hittite	no inherited word

Shared vocabulary dates a stage of the language only if the argument is sound: this one is debated.

Tier B. Sources: Lazaridis et al. 2022 (supplied); Heggarty et al. 2023 and Kassian and Starostin 2023 (supplied). Forms to verify (Fortson; Adams).

Pastoral words: cow, sheep, horse

cow

*g^{w}ṓus · nearly every branch

Old Irish	**bó**
Latin	**bōs**
Greek	**boûs**
Armenian	**kov**
Avestan	**gāuš**
Sanskrit	**gáu-**
Tocharian B	**keu**
Luwian	**wawa-**

sheep

*h$_2$ówis · 11 groups

Latin	**ovis**
Greek	**óis**
Sanskrit	**ávi-**
Luwian	**hāwa/i-**
Tocharian B	**āu “ewe”**
Lithuanian	**avìs**
Old Irish	**ói**
English	**ewe**

horse

*h$_1$ékwos · all but Slavic

Old Irish	**ech**
Latin	**equus**
Old English	**eoh**
Greek	**híppos**
Avestan	**aspa-**
Sanskrit	**áśva-**
Tocharian B	**yakwe**
Luwian	**azu(wa)-**

***péku “livestock”** (domestic animals only, as against wild *g^{wh}ēr): Latin pecu, Sanskrit paśu-, Avestan pasu, Lithuanian pekus. The authors call the **horse** “certainly known” in PIE, but its domestic status arguable.

Tier A: Mallory and Adams 2006, ch. 9.2, pp. 134–142 (supplied; paraphrased).

A poetic formula that travels

Greek (Iliad 9.413)

kléos áphthiton

"imperishable fame"

Sanskrit (Rigveda)

śrávaḥ ákṣitam

"imperishable fame"

PIE *Ḱléwos ńdʰgʷʰitom: the same phrase in two languages 2,000 miles apart (Kuhn 1853; Watkins 1995).

Poetry as inherited vocabulary: formulas survive when single words change.

Tier A for phrases (Kuhn 1853; Watkins 1995; Drive library O14). Rigveda locus to verify (Jamison and Brereton 2014, N01).

A poetic formula: “he slew the serpent”

*(h₁e)gʷʰént h₁ógʷʰim “he slew the serpent”

Avestan

janat azīm

“he struck the serpent”

Vedic Sanskrit

áhann áhim

“he slew the serpent”

Greek

kteíne óphin

“he killed the snake”

Hittite

illuyanka kwenta

“he killed Illuyanka”

Dragon-slaying myth

Indra and the serpent Vṛtra in the Rigveda, and the Hittite storm-god and Illuyanka, tell a similar story. The shared phrase is a clue to inherited poetry.

A caution

The words match closely only in Indo-Iranian. Reconstruction rarely reaches beyond two-word phrases, so treat this as a debated example (tier C).

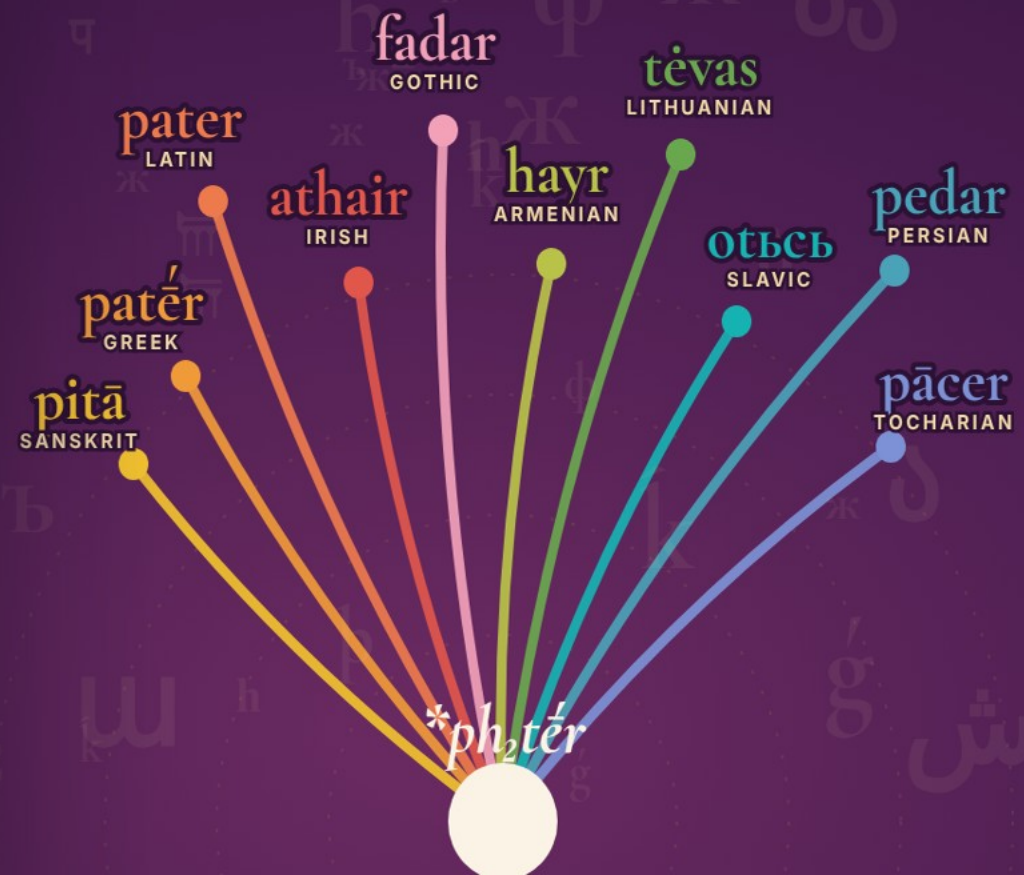
Tier C: Mallory and Adams 2006, ch. 7.8, p. 117 (supplied; paraphrased).

PART 3

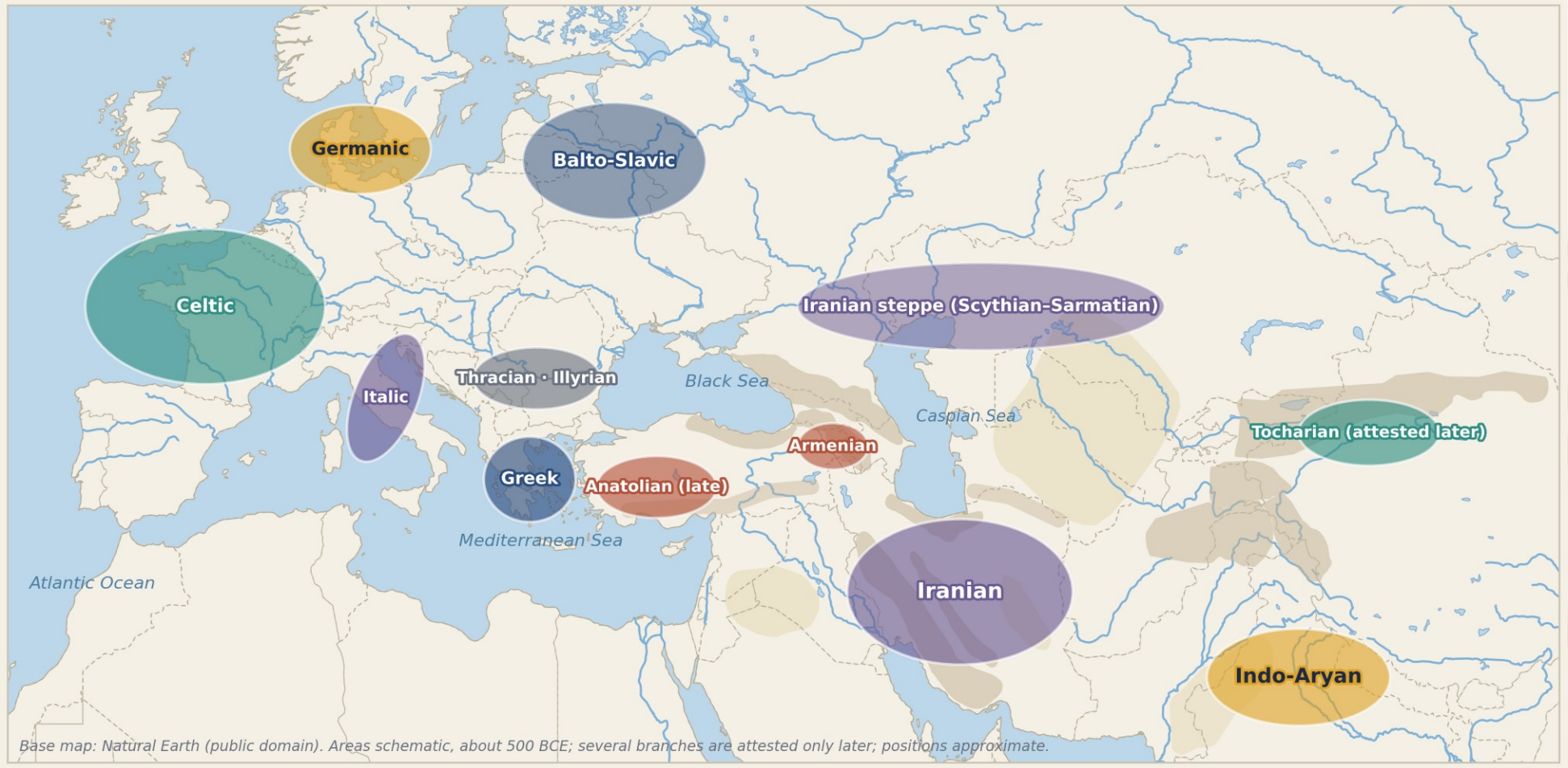
03

The family

Ten branches, their oldest records and their neighbours



The branches around 500 BCE (schematic)



Base map: Natural Earth (public domain). Language areas drawn for this course; approximate; many branches attested only later.

The ten branches

Anatolian
extinct

Hittite, Luwian, Lycian

OLDEST RECORD
Hittite tablets, c. 1650 BCE

Tocharian
extinct

Tocharian A and B

OLDEST RECORD
Tarim Basin, 6th–8th c. CE

Greek
living

Mycenaean to Modern Greek

OLDEST RECORD
Linear B, c. 1450–1200 BCE

Armenian
living

Classical and Modern

OLDEST RECORD
5th century CE

Albanian
living

Gheg, Tosk

OLDEST RECORD
15th century CE

Indo-Iranian
living

Sanskrit, Avestan, Persian, Hindi, Pashto

OLDEST RECORD
Old Persian, c. 520 BCE

Balto-Slavic
living

Lithuanian, Russian, Polish

OLDEST RECORD
Church Slavonic, 9th c. CE

Germanic
living

English, German, Swedish

OLDEST RECORD
Runes, 2nd century CE

Italic
living

Latin, Spanish, French, Italian

OLDEST RECORD
Latin, 6th century BCE

Celtic
living

Irish, Welsh, Breton

OLDEST RECORD
Ogham, 4th century CE

Grouping: Heggarty et al. 2023 (supplied). Member lists standard (tier A); verify in Fortson.

How the branches are grouped

Anatolian

Usually first to separate; possibly a sister of PIE (Indo-Hittite).

Indo-Iranian

The only uncontroversial subgroup.

Balto-Slavic

A majority view, disputed by some Balticists.

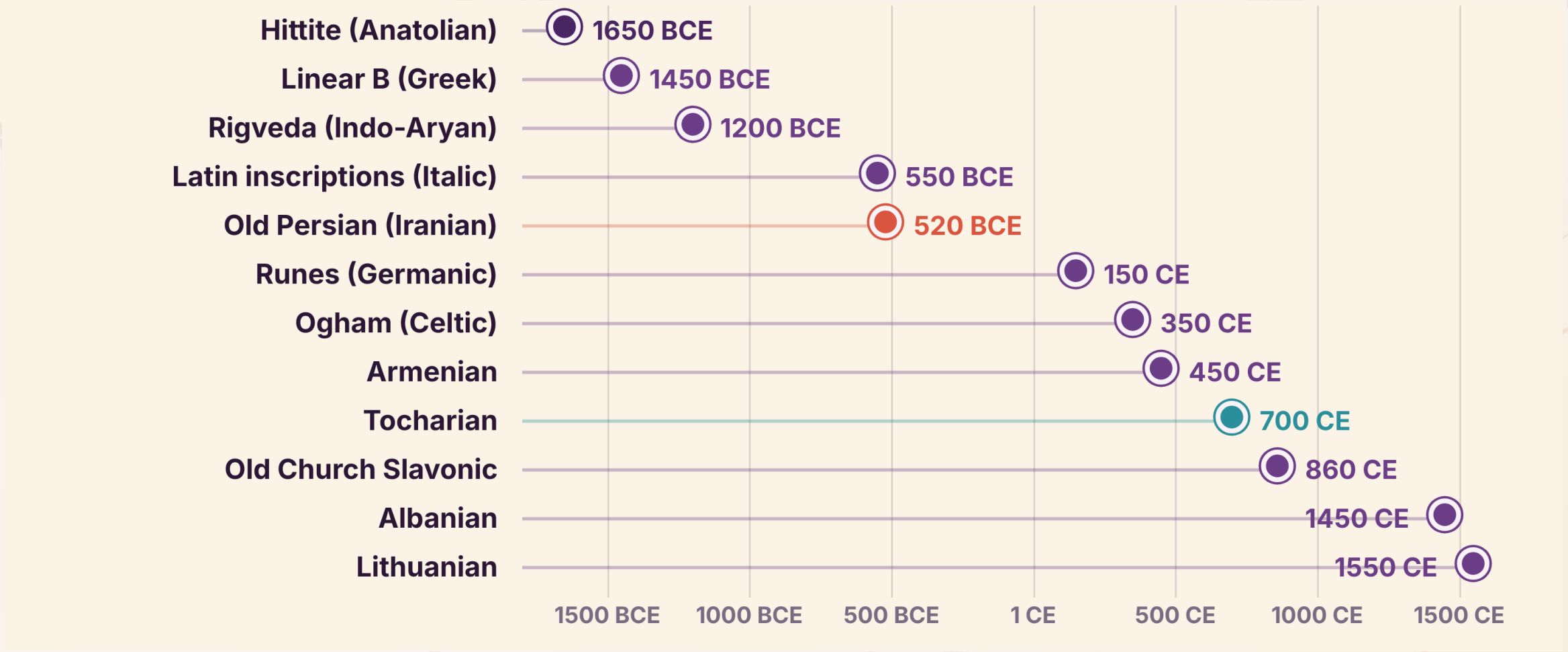
Tocharian

No firm link to any branch; probably an early split in the east.

Subgrouping rests on shared innovations: some groups are firm, the deeper order is open.

Tier A for the positions as reported; C for the Indo-Hittite question. Mallory and Adams 2006, ch. 5, pp. 73–84 (supplied).

Oldest written records, on a timeline



Tier A/B, standard; verify in Fortson / Mallory and Adams. Behistun date: Luschey, Encyclopaedia Iranica (supplied via Drive).

Anatolian and Tocharian: the odd ones out

Anatolian

The oldest written Indo-European languages (Hittite, from about 1650 BCE). They keep archaic features, which is why some scholars place them first to split.

Tocharian

Two languages of the Tarim Basin, known from manuscripts of the 6th–8th centuries CE. Indo-European speakers in western China.

Why they matter

They test the family: any model of the homeland must explain a branch in Anatolia and one in the Tarim Basin.

The order of the deepest splits is weakly supported (Heggarty et al. 2023).

Sources: Heggarty et al. 2023; Lazaridis et al. 2025 (supplied). Features: standard, verify in Kloekhorst; Melchert; Adams.

Why Anatolian looks archaic

What Anatolian keeps

The laryngeal *h_2
Hittite ḫant- “front”, from the same root as Greek antí and Latin ante

Heteroclitic r/n nouns
Hittite wātar “water”, genitive witenas

Two verb conjugations
the -mi and the -ḫi type

What Anatolian lacks

No dual number
only singular and plural

No subjunctive or optative
the moods of other branches are absent

Doubtful feminine, no augment
a simpler grammatical system

Two readings. Indo-Hittite: Anatolian is a sister of PIE, so it split off first and the “missing” features never existed. Anatolian-as-daughter: it lost them. Either way Anatolian is generally seen as the first to separate, and the 2022 and 2025 DNA papers use “Indo-Anatolian” for the common ancestor of both. The Indo-Iranian branch is the only *uncontroversial* subgroup.

Tier A/C: Mallory and Adams 2006, ch. 5, pp. 73–75 (supplied; paraphrased).

Greek, Armenian, Albanian: three isolates

Greek

Linear B (c. 1450 BCE) onwards: the longest continuous written record in the family.

Armenian

Written from the 5th century CE; full of Iranian loanwords (including month names such as Mehekan and Trē).

Albanian

First texts in the 15th century CE; a single branch with many Latin, Greek and Slavic loans.

A special Greek–Armenian relationship is proposed but disputed.

Tier A for main facts; Armenian month names: Panaino, Calendars i (supplied via Drive); Graeco-Armenian link tier C.

Italic, Celtic, Germanic, Balto-Slavic

Italic

Latin, Oscan, Umbrian. Latin gave the Romance languages.

Celtic

Gaulish and Celtiberian (extinct); Irish, Welsh, Breton, Scottish Gaelic.

Germanic

Gothic (extinct); English, German, Dutch, Scandinavian languages.

Balto-Slavic

Lithuanian, Latvian; Russian, Polish, Czech, Serbian and other Slavic languages.

Whether Italic and Celtic form one unit ("Italo-Celtic") is still debated.

Tier A for membership; Italo-Celtic: debated. Balto-Slavic and Indo-Iranian shared innovations: Kassian and Starostin 2023 (supplied).

Lost and thinly known languages

Language	Where	Record
Phrygian	Central Anatolia	short inscriptions
Thracian, Dacian	Balkans	names, glosses, a few inscriptions
Illyrian, Messapic	Adriatic coasts	names; Messapic inscriptions
Venetic	Northern Italy	short inscriptions
Lusitanian, Celtiberian	Iberia	few long inscriptions
Gaulish	Gaul	inscriptions and place names
Sogdian, Bactrian, Khotanese	Central Asia	manuscripts (Iranian)
Scythian, Sarmatian	Steppe	names and glosses (Iranian)

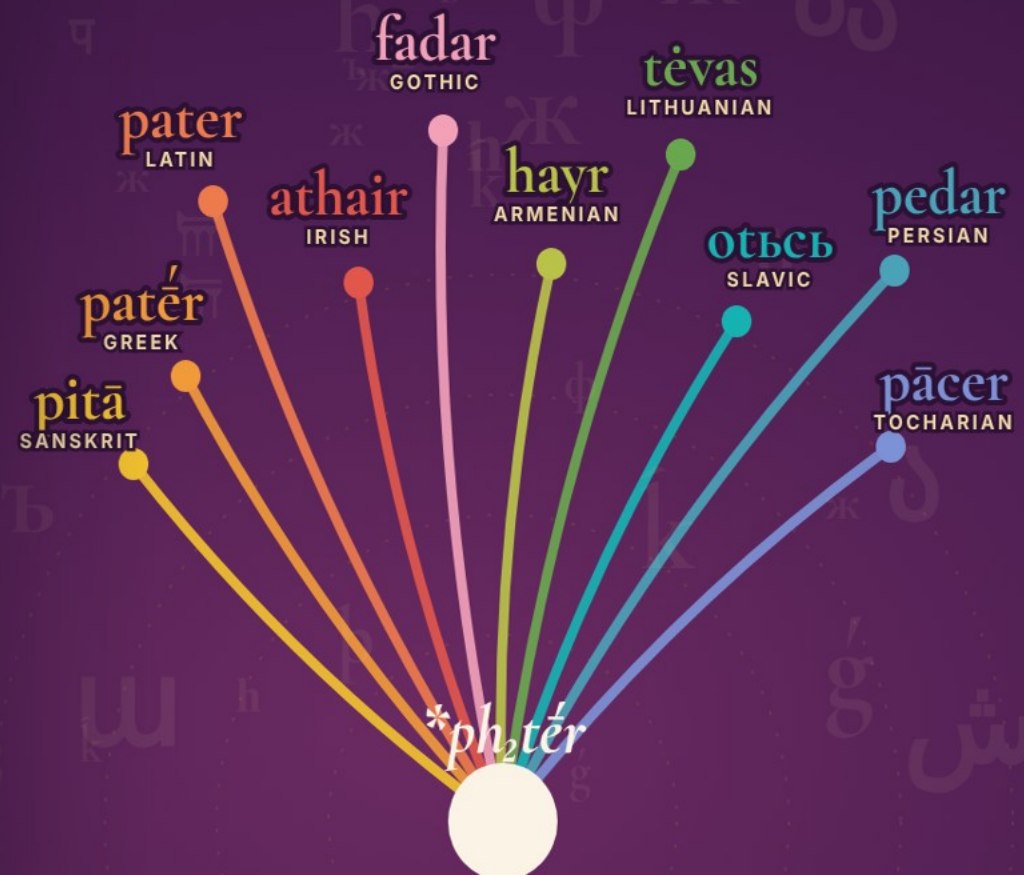
Tier B; verify in Fortson, Clackson and Windfuhr 2009 (Drive library K01, K07, M01).

PART 4

04

Trees and dates

What makes a branch, and how old is the root?



Tree and wave

Tree (Schleicher)

Languages split and diverge, like branches. Good for describing sequences of splits.

Wave (Schmidt)

Innovations spread across neighbouring dialects. Good for explaining overlapping features.

In practice

Scholars use both: trees for splits, waves and contact for overlaps such as satem/centum.

Tier A, standard history. Heggarty et al. 2023; Kassian and Starostin 2023 (supplied).

Shared innovations make a branch

Group	Shared innovation	Status
Indo-Iranian	a-merger; RUKI rule; genitive plural *-nām	secure (Schmitt 1987)
Germanic	Grimm's Law; Verner's Law	secure
Balto-Slavic	shared accent and vocabulary changes	usually accepted
Italo-Celtic	some shared endings	debated
Graeco-Armenian	some shared vocabulary and morphology	debated
Deepest order	Anatolian, Tocharian, others	weakly supported (<26%)

Sources: Schmitt 1987 (Indo-Iranian); Heggarty et al. 2023 (supplied). Others: standard, tier B/C.

Borrowing: what to trust when tracing descent

Layer of language	How easily borrowed	Note or example
Vocabulary	Most easily	Romani: loans from Greek, Persian, Armenian
Sound system	Resists	Phonology is a better guide to descent than word lists
Derivational affixes	Less often	e.g. the suffix -nik
Inflection, syntax	Least often	Syntax usually follows lexical loans (Domari numerals from Arabic)

Descent is best shown by sound and grammar, with vocabulary as the layer to check for loans.

Tier A for the hierarchy as presented; tier C for the critique of Bayesian trees. Pereltsvaig and Lewis 2015, pp. 60–66 (supplied, partial; polemical). Ranking is the book's summary of earlier work.

Why glottochronology was dropped

Branch	Dates from the two word-lists (bc)	Gap
Hittite	6400 or 8800	2,400 years
Armenian	5700 or 4700	1,000 years
Sanskrit	3700 or 2900	800 years

- Two lists of different length gave different dates for the same split.
- No constant core list and no constant rate: English kept about 68%, Icelandic about 97%.

Tier A: Mallory and Adams 2006, p. 95, Table 6.2 (Tischler 1973; supplied). Gaps computed from the two dates. Modern Bayesian dating is a different method (next slides).

How a date is estimated

1

Code cognates

IE-CoR dataset: 161 languages

2

Model change

rates of word replacement

3

Calibrate

known dates of old texts

4

Sample trees

Bayesian computation

5

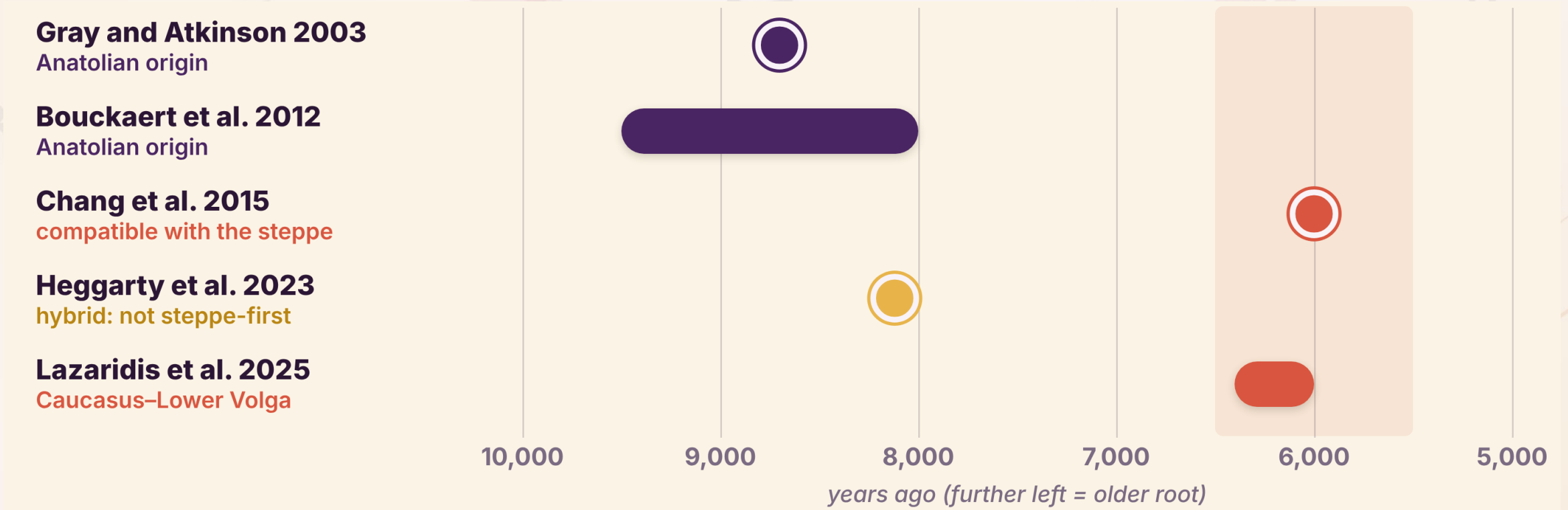
Report

date with an interval

A date is a statistical estimate with a wide interval, not a measurement.

Sources: Heggarty et al. 2023 (supplied); Kassian and Starostin 2023 (supplied). Method summary: standard.

Estimates of the root date



The estimates differ by thousands of years: the models and their assumptions drive the answer. Lazaridis 2025 is converted from 4400–4000 BCE by adding about 2,000 years.

Heggarty 2023 and Lazaridis 2025: verified (supplied). Gray–Atkinson 2003, Bouckaert 2012, Chang 2015: tier B, verify in the originals (Drive Q22, Q23).

Dating PIE from texts, words and things

Texts

Earliest written records: Anatolian about 2000 bc, Indo-Iranian about 1400 bc, Greek about 1300 bc or earlier.

Rate of change

Hittite to Sanskrit shows 1,500 to 2,000 years of change: about 3000 bc, plus or minus a millennium.

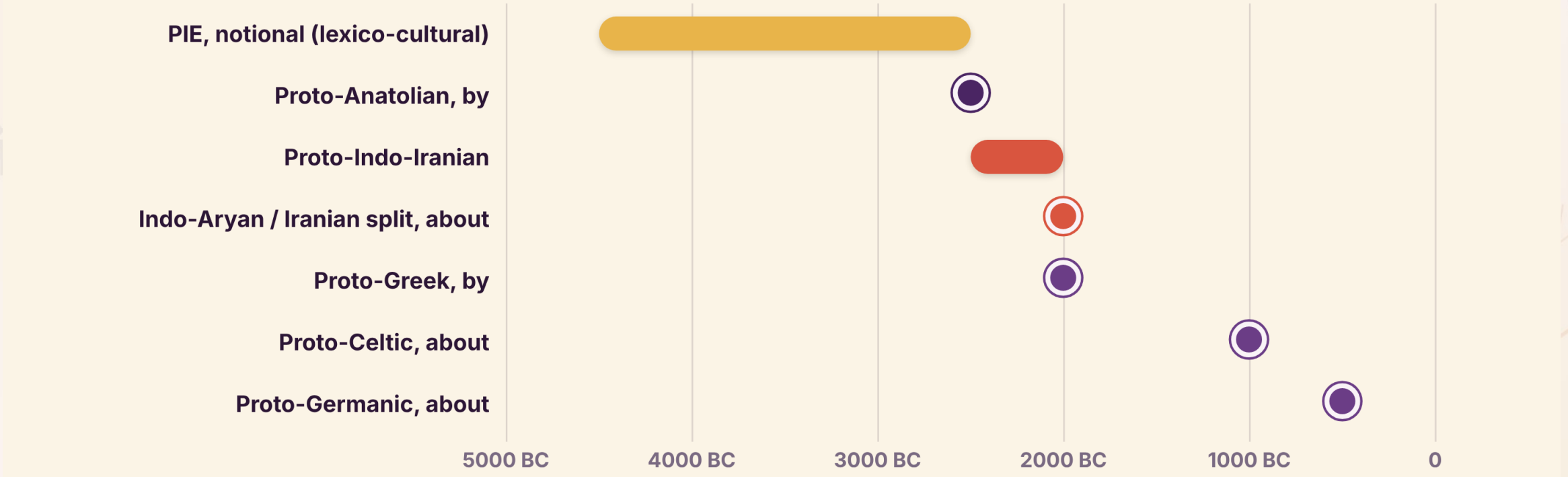
Words and things

Neolithic words, plus wheel, wool and plough words: about 4500 to 2500 bc.

Different methods overlap in the fourth to third millennium bc: a notional range, not a measurement.

Tier A (authors' stated estimates) with tier C for the equations used. Mallory and Adams 2006, ch. 6, pp. 87–103 (supplied; paraphrased). Compare the Bayesian dates in slides 35 and 37.

A chronology of the daughter languages



Reading: the authors place PIE at about 4500–2500 BC, because it has Neolithic words (cattle, sheep, plough) and wheeled-vehicle vocabulary, and the earliest wheels are fourth-millennium. Groups documented in the Bronze Age get deep dates while western and northern groups are dated only from their first records. Zimmer's caution: avoid "PIE" for anything older than c. 2500 BC.

Tier A (authors' working dates): Mallory and Adams 2006, ch. 6, pp. 86–104 (supplied; paraphrased). Compare the statistical dates on slide 35.

Three misreadings to avoid

THE MISREADING

“These words look alike, so the languages are related.”

THE BETTER READING

Resemblance is weak evidence. Only sound correspondences that recur across many basic words count; loans and chance are screened out first.

THE MISREADING

“A starred form is an attested word.”

THE BETTER READING

An asterisk marks a hypothesis. Only 1% of 1,474 reconstructions have cognates in all twelve groups; most rest on six or fewer.

THE MISREADING

“A date for the root is a measurement.”

THE BETTER READING

It is a model estimate with a wide interval, and published ones differ by thousands of years.

Tier A: Mallory and Adams 2006; Heggarty et al. 2023. Synthesis by Claude.

Five questions for any “related languages” claim

1. **Which words?**
   Basic vocabulary: body parts, numerals, kinship, not culture words.

2. **Which correspondences?**
   Do the sounds match regularly in many words?

3. **Loan or cognate?**
   Could one language have copied the word from the other?

4. **Which innovation?**
   Shared changes define a branch, not shared old features.

5. **Which date, how wide?**
   Look for the interval and for the model behind it.

Synthesis by Claude from slides 5, 6, 31 and 34.

The dispute over the 2023 trees

Secure

The ten main branches and shared innovations within them.

Contested

The root date, the order of the deepest splits, and the homeland.

Why it matters

Language dates must agree with archaeology and DNA before a homeland model is accepted.

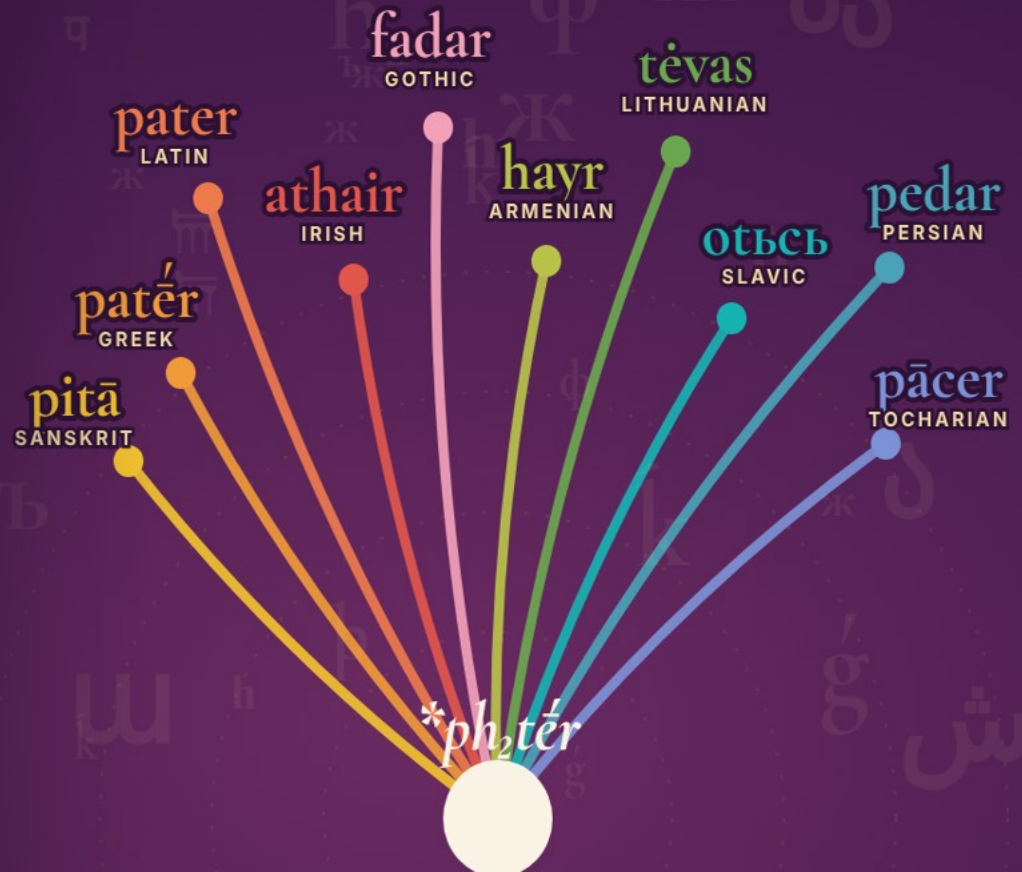
Sources: Kassian and Starostin 2023; Heggarty 2025; Heggarty et al. 2023 (supplied).

PART 5

05

Language, genes and Persian

What each kind of evidence can and cannot show



Genes are not languages

DNA shows people

Ancient DNA shows movements and mixing of populations, not the language they spoke (Haak et al. 2015).

Languages shift

Languages can be replaced without large migrations: Latin in Gaul, Arabic in Egypt.

Three lines

A strong model needs linguistics, archaeology and DNA to agree.

The 2025 Indo-European study treats DNA as one line of evidence, not proof of a language.

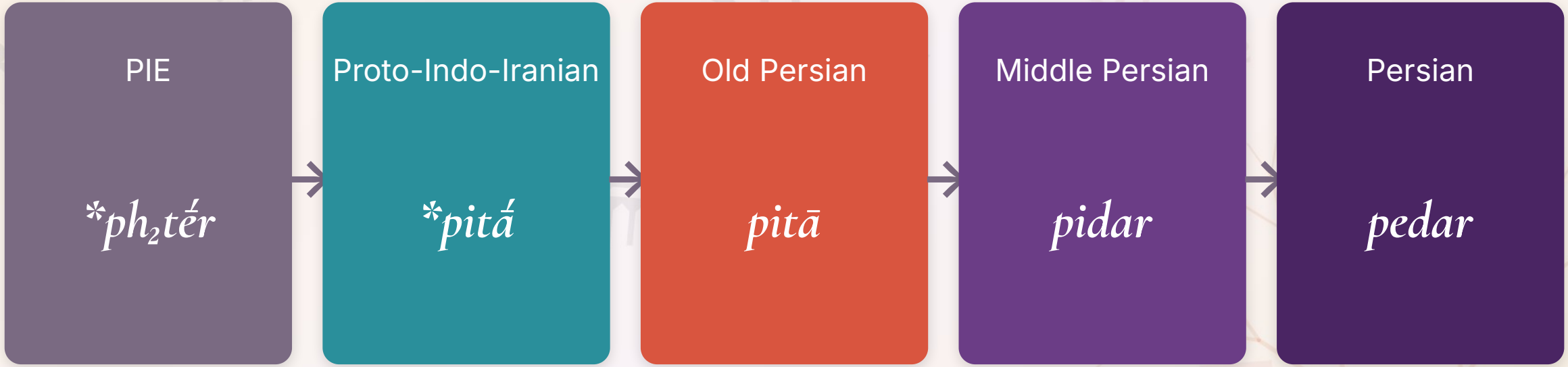
Source: Haak et al. 2015 (supplied). Language-shift examples: standard.

Borrowing between families

Case	Evidence	Status
Finnish sata "100"	matches Indo-Iranian *śata- (Persian sad)	standard (tier B)
Indo-Iranian substratum	words without IE etymology, linked to the BMAC language	Lubotsky 2001 (tier B)
"wine"	Hittite wiyana, Greek oinos, Latin vinum; also in Kartvelian and Semitic	Wanderwort (tier A)
"paradise"	Old Iranian *pari-daiza- into Greek, Latin, English	Module 5 (tier A)

Tier B; verify in Lubotsky 2001 (Drive M19), Kloekhorst 2008, Fortson. Loans in Persian and English: Module 5.

Where Persian sits in the family



Module 5 follows this chain for dozens of words.

*Verified chain: Old Persian pitā, Middle Persian pidar (Wiktionary etymology lines via Drive). Proto-Indo-Iranian *pitā́: standard; verify.*

Check your understanding

1. What separates a cognate from a loan and from chance?
2. What does Grimm’s Law say about PIE *p in Germanic? Give a Latin/English pair.
3. What did Saussure predict in 1879, and what confirmed it in 1927?
4. What share of 1,474 reconstructions have cognates in all twelve groups?
5. How does Verner’s Law explain the Sanskrit pitā́ / Gothic fadar pair?
6. Which branch has the oldest written record, and about when?
7. Why was glottochronology dropped?
8. Why do root-date estimates of 8,700 and 6,000 years ago differ so much?
9. What does the centum / satem split show, and is it a family tree?
10. Name the four kin words with cognates across branches on the family-words slide.

Questions drawn from slides 5 to 35.

Check your understanding: answers

1. **Regular recurrence**
   A cognate recurs in many words with regular sound shifts; a loan is copied and irregular; chance does not recur.

2. **p → f**
   Latin pēs / English foot.

3. **Laryngeals**
   Extra PIE sounds; Kuryłowicz read Hittite ḫ as one of them.

4. **1%**
   75% rest on cognates in six groups or fewer.

5. **Accent**
   Accent after t gives voiced d (fadar); accent before gives þ (brōþar).

6. **Hittite, c. 1650 BCE**
   Anatolian cuneiform tablets.

7. **No constant rate**
   Different lists gave different dates; English kept about 68% and Icelandic about 97%.

8. **Models drive the answer**
   Anatolian-origin studies date near 8,700; steppe-compatible near 6,000.

9. **A sound pattern**
   *ḱ became k/h in centum and s/š in satem languages; it is areal, not a tree.

10. **Father, brother, sister, daughter**
    *ph₂tḗr, *bʰréh₂tēr, *swésōr, *dʰugh₂tḗr.

Tier A: Mallory and Adams 2006; Heggarty et al. 2023; Fortson.

Key terms

Cognate
A word that descends from the same ancestor word in related languages: pedar and father.

Sound law
A regular sound change, recurring across many words in one language.

Proto-language
The reconstructed ancestor of a family; its forms carry an asterisk.

Reconstruction
A hypothesis about an unwritten ancestor form, built from cognates.

Laryngeal
A PIE sound predicted by Saussure and later seen in Hittite ḫ.

Labiovelar
A k-sound with lip rounding (*kʷ); it splits into P and Q forms.

Branch
A group of languages sharing innovations from one ancestor.

Shared innovation
A change that arose once and is inherited; it defines a branch.

Loanword
A word copied from another language, usually irregular in sound.

Glottochronology
A dated-by-word-list method, dropped for uneven rates of change.

Bayesian dating
Statistical models that sample trees and report a date with an interval.

Subgrouping
Placing branches on a tree by their shared innovations; the deepest order is debated.

Definitions summarised from the module slides and standard handbooks (Fortson; Clackson).

More key terms

Centum / satem
Two outcomes of the palatovelar *ḱ: k or h versus s, š or ś.

Asterisk (*)
Marks a reconstructed form no one recorded.

Heteroclitic
A noun whose stem alternates, as Hittite wātar / witenas.

Indo-Hittite
The view that Anatolian is a sister of PIE, not a daughter.

Aspirate
A stop pronounced with a breath of air, such as b^h; PIE had voiced aspirates.

Palatovelar
A k-type sound at the palate; ḱ, ǵ and $ǵ^h$.

Reflex
The descendant of a proto-sound or word in a daughter language.

Substrate
A language replaced by another, leaving loanwords in it.

Lexicon
The vocabulary of a language; basic words are the best witnesses.

Isogloss
A line on a map around the area where a feature occurs.

Neogrammarians
1870s school: sound laws admit no exceptions.

Cognate set
All the related words that go back to one proto-word.

Definitions summarised from the module slides and standard handbooks.

Takeaways

1

Sound laws are the proof

Regular correspondences, with explained exceptions, are what make a family a family.

2

Ten secure branches

The branches and their shared innovations are secure; the deepest order is not.

3

Dates are estimates

Root dates range over thousands of years; models, not measurements, drive the answer.

Next: Module 2 (archaeology of the steppe) and Module 3 (ancient DNA).

Key literature

- Jones, W. (1788 [1786]). The Third Anniversary Discourse. Asiatick Researches 1.
- Fortson, B. (2010/2019). Indo-European Language and Culture. Wiley-Blackwell (to obtain).
- Mallory, J. P. and Adams, D. Q. (2006). The Oxford Introduction to PIE and the PIE World (supplied; ch. 3, 5–7, 9 read).
- Clackson, J. (2007). Indo-European Linguistics: An Introduction. CUP (to obtain).
- Beekes, R. (2011). Comparative Indo-European Linguistics. Benjamins (to obtain).
- Watkins, C. (1995). How to Kill a Dragon. OUP (to obtain).
- Schmitt, R. (1987). "Aryans". Encyclopaedia Iranica (supplied).
- Haak, W. et al. (2015). Nature 522: 207–211 (supplied).
- Heggarty, P. et al. (2023). Science 381: eabg0818 (supplied).
- Kassian, A. and Starostin, G. (2023). Informal review of Heggarty et al. (supplied).
- Heggarty, P. (2025). Commentary on Lazaridis et al. 2025 (supplied).
- Lazaridis, I. et al. (2022). Science 377: eabm4247 (supplied).
- Lazaridis, I. et al. (2025). Nature 639: 132–142 (supplied).
- Gray, R. and Atkinson, Q. (2003). Nature 426: 435–439 (to obtain).
- Bouckaert, R. et al. (2012). Science 337: 957–960 (to obtain).
- Chang, W. et al. (2015). Language 91: 194–244 (to obtain).
- Pereltsvaig, A. and Lewis, M. W. (2015). The Indo-European Controversy. CUP (supplied; Introduction–ch. 4 start read).

Citations compiled from the project library and standard works; verify details before citing.

Sources and credits for this module

- Map: Natural Earth (public domain) base with schematic overlays drawn for this course.
- Quotation: William Jones (1786), public domain.
- Grouping, dates and critiques: Heggarty et al. 2023; Kassian and Starostin 2023; Heggarty 2025; Lazaridis et al. 2022, 2025 (supplied papers).
- Indo-Iranian innovations, Behistun date, scripts, Armenian month names: Schmitt 1987; Luschey (Bisotun ii); Skjærvø (Iran vi 3); Panaino (Calendars i), Encyclopaedia Iranica (Drive).
- Standard comparative forms (numerals, kinship, sound laws): tier A/B; to be verified in Fortson, Beekes, Mallory and Adams, de Vaan and Kloekhorst once obtained.
- Added in v2: reliability, labelling rules, subgrouping, glottochronology and dating of PIE from Mallory and Adams 2006 (ch. 5–7); borrowing hierarchy from Pereltsvaig and Lewis 2015 (pp. 60–66). Mallory 1989 and the later chapters of both books were unreadable and are not used.