Alexei S. Kassian (kassian@iling-ran.ru), Mikhail Zhivlov, George Starostin, Artem A. Trofimov, Petr A. Kocharov, Anna Kuritsyna and Mikhail N. Saenko

Text supplement: trees & comments

MrBayes commands

The following commands for the MrBayes package were applied:

```
FORMAT DATATYPE=RESTRICTION GAP=- MISSING=?

lset coding=noabsencesites rates=gamma covarion=yes;

calibrate Old_Hittite = uniform(3500, 3650);
calibrate Tocharian_B = uniform(1100, 1600);
calibrate Ancient_Attic_Greek = fixed(2375);
calibrate Classical_Armenian = uniform(1500, 1600);
calibrate Archaic_Latin = fixed(2200);
calibrate Old_Irish = uniform(1100, 1300);
calibrate Proto_Brittonic = uniform(1400, 1700);
calibrate Proto_Germanic = uniform(2300, 2500);
calibrate Proto_Slavic = uniform(1700, 2000);
calibrate Proto_East_Baltic = uniform(2000, 2400);
calibrate Old_Indic_Atharvaveda = uniform(3000, 3200);
calibrate Proto_Iranian = uniform(3000, 3500);
calibrate Albanian = fixed(50);

prset clockratepr=exponential(3e5);

prset speciationpr=exp(1);

prset extinctionpr=beta(1,1) nodeagepr=calibrated;

prset brlenspr=clock:fossilization clockvarpr=TK02;

prset treeagepr = uniform(5500,10500);

prset samplestrat=fossiltip;

showmodel;
mcmcp ngen=10000000 printfreq=10000 samplefreq=500 nruns=4
nchains=4 savebrlens=yes;
mcmc;
sumt relburnin=yes burninfrac=0.25;
sump relburnin=yes burninfrac=0.25;
```

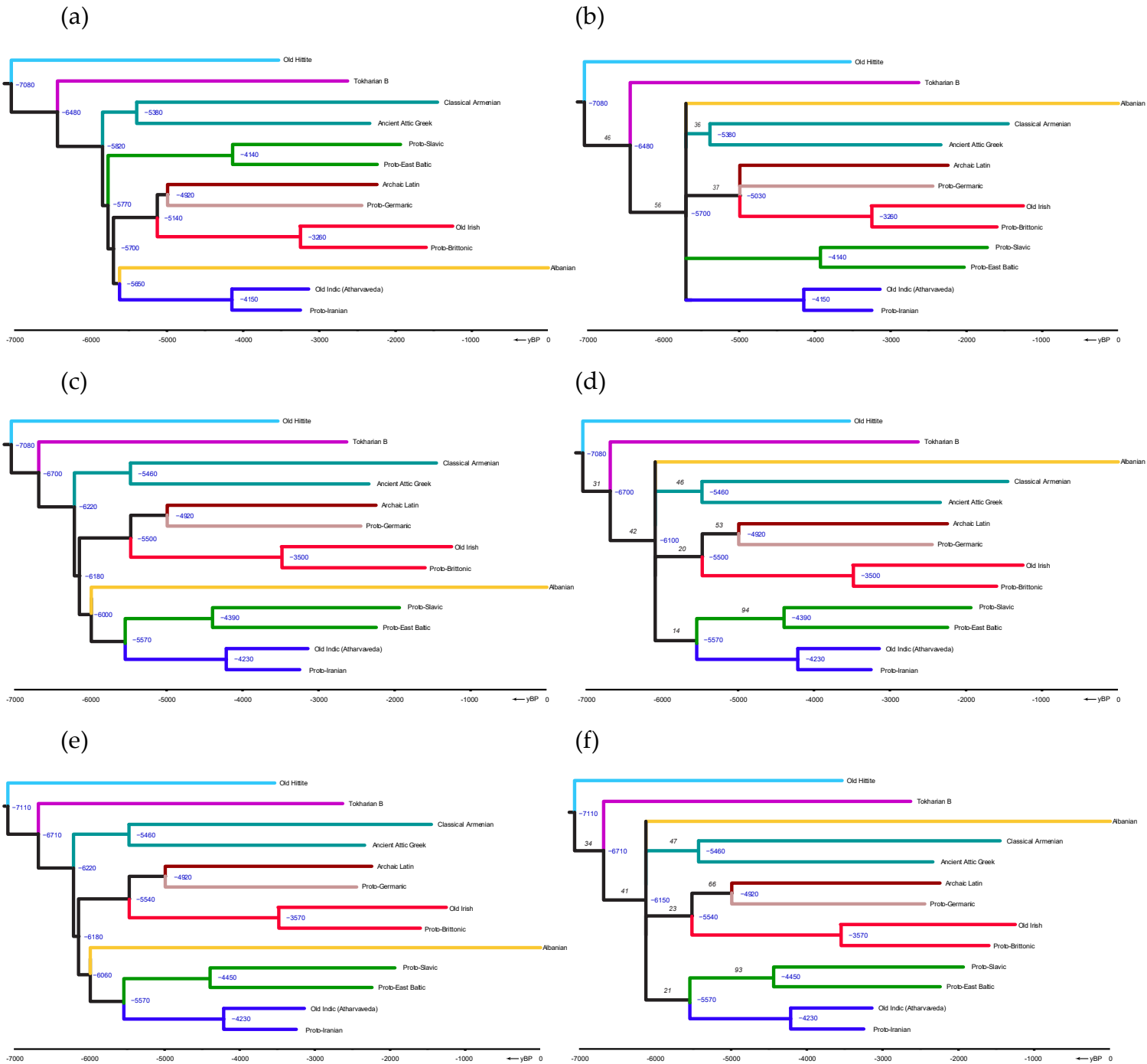# Phylogenetic trees obtained by individual algorithms

Fig. S1. Phylogenetic trees of the Indo-European language family produced by the **StarlingNJ** method from the multi-state matrix. Bootstrap values for the trees (b), (d), (f) are shown in italic near the branches (not shown for stable nodes with bootstrap value ≥ 95%). The trees are dated. Divergence times are given to the right of each node. Datasets with and without the Proto-Samoyed outgroup produce identical topologies and dates, thus only the trees for the proper Indo-European dataset are offered. Traditional subgroups are identified by color branches.

- **Stage-1 dataset with root cognacy** (*wind = veter*, *agni = ignis*): **(a)** binary nodes; **(b)** neighboring nodes are joined if the distance between them is ≤ 300 years.
- **Stage-2 dataset without derivational drift** (*wind ≠ veter*, *agni = ignis*): **(c)** binary nodes; **(d)** neighboring nodes are joined if the distance between them is ≤ 300 years.
- **Stage-3 homoplasy-optimized dataset** (*wind ≠ veter*, *agni ≠ ignis*): **(e)** binary nodes; **(f)** neighboring nodes are joined if the distance between them is ≤ 300 years.

For Stage-2 and Stage-3, no topological discrepancies between the trees, i.e., within the pairs (c) & (e), (d) & (f).
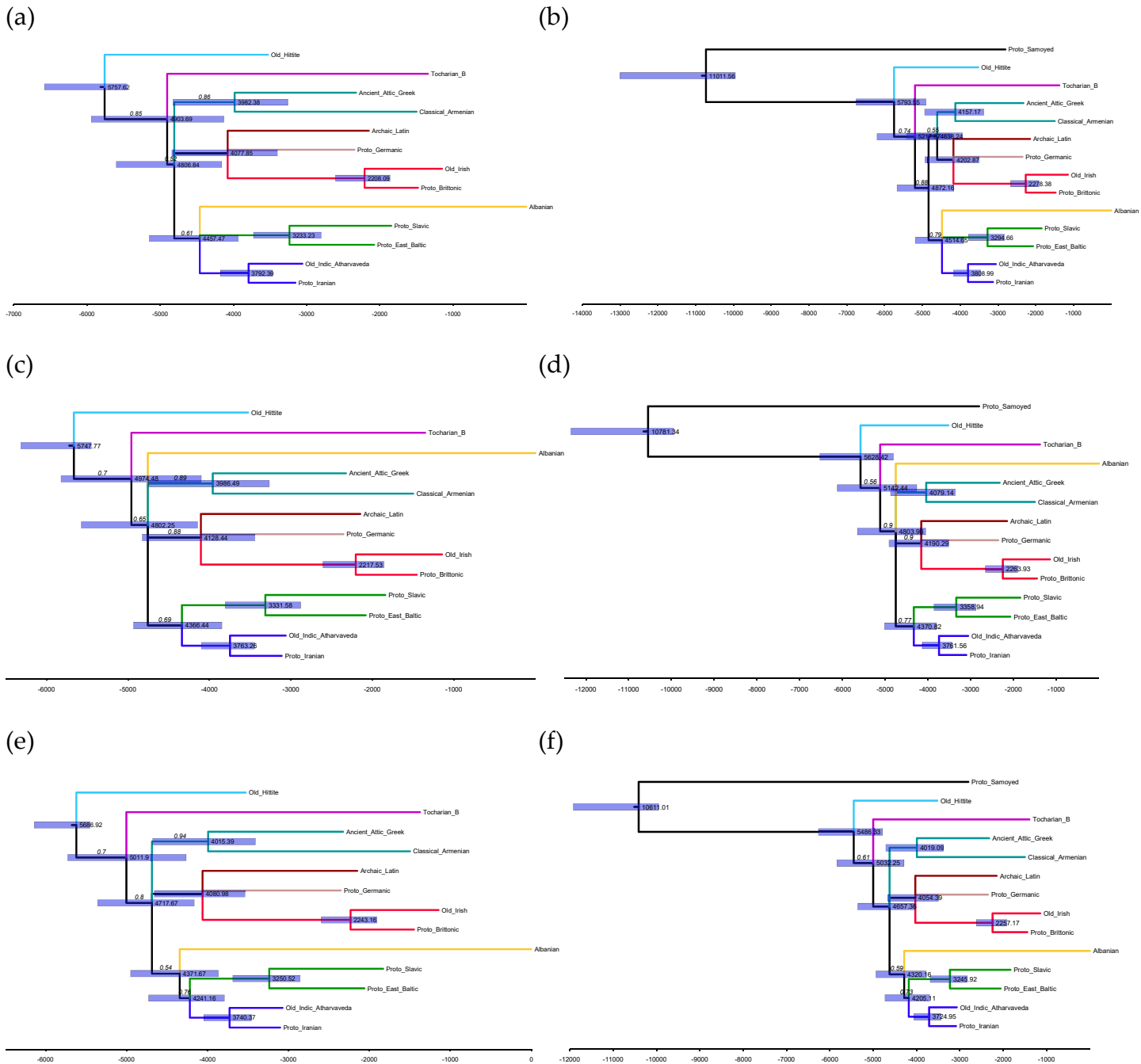
Fig. S2. Phylogenetic trees of the Indo-European language family produced by the Bayesian MCMC method from the binary matrix in the MrBayes software (50% majority rule trees). No topological constraints. No chronological constraints for intermediate nodes; root is predefined within the range 10,500–5,500 yBP (for the proper IE dataset) and with the upper limit 10,000 yBP and mean 20,000 yBP (for the IE-Samoyed dataset). Bayesian posterior probabilities are shown in italic near the branches (not shown for stable branches with P ≥ 0.95). Blue bars represent the 95% highest probability density (HPD) for the divergence times; mean divergence times are given to the right of each node. Scale values represent years before present (yBP). Traditional subgroups are identified by color branches.

- **Stage-1 dataset with root cognacy** (*wind = veter*, *agni = ignis*): **(a)** proper IE, **(b)** IE-Samoyed.
- **Stage-2 dataset without derivational drift** (*wind ≠ veter*, *agni = ignis*): **(c)** proper IE, **(d)** IE-Samoyed.
- **Stage-3 homoplasy-optimized dataset** (*wind ≠ veter*, *agni ≠ ignis*): **(e)** proper IE, **(f)** IE-Samoyed.

For Stage-2 and Stage-3, no topological discrepancies with and without Proto-Samoyed, i.e., within the pairs (c) & (d), (e) & (f). For Stage-1, topological discrepancies between (a) & (b) are insignificant.
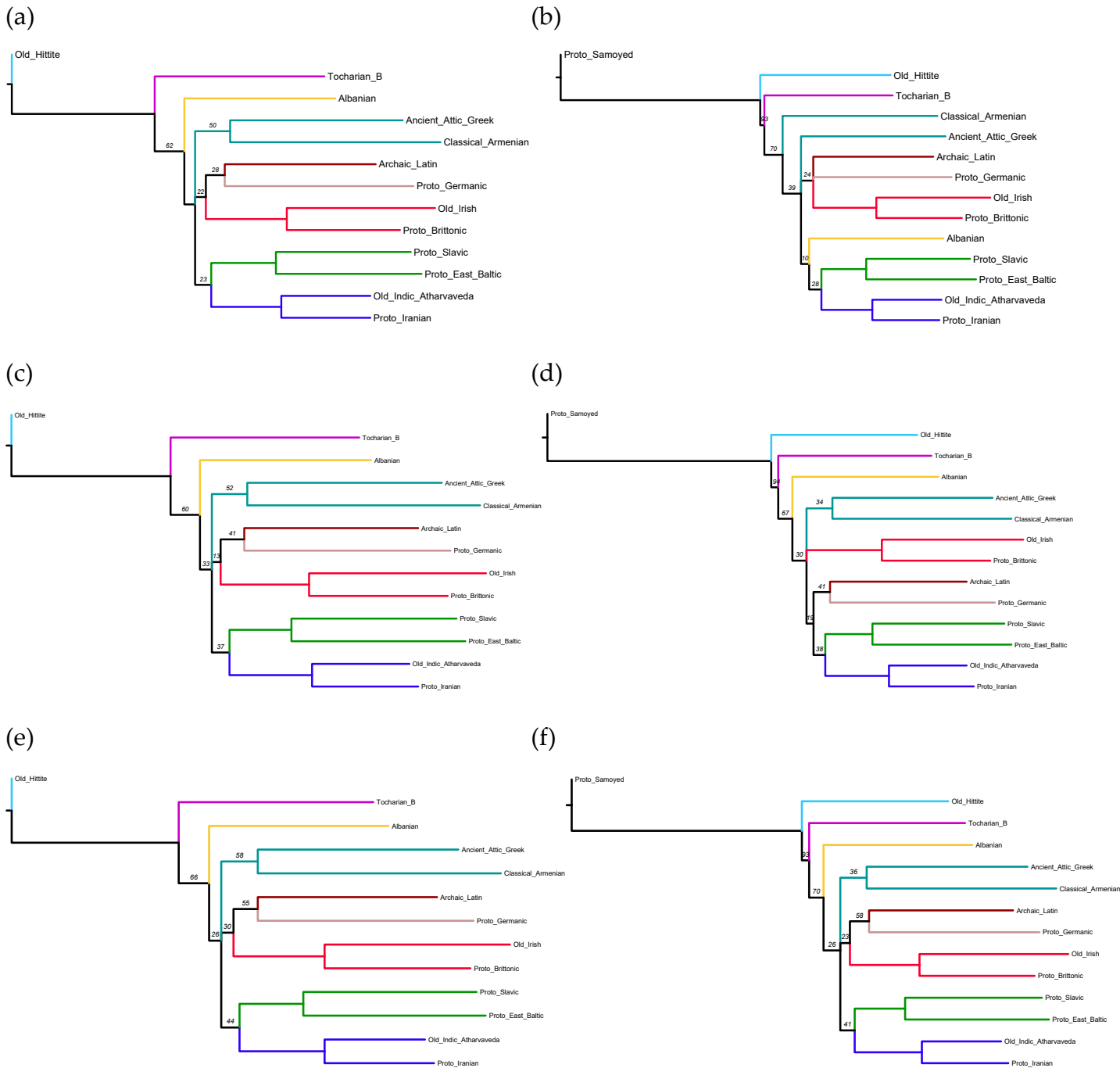
4

Fig. S3. Majority-rule consensus phylogenetic trees of the Indo-European language family produced by the Unweighted Maximum Parsimony (MP) method from the binary matrix in the TNT software. Bootstrap values are shown in italic near the nodes (not shown for stable nodes with bootstrap value ≥ 95%). Branch length reflects the relative rate of cognate replacement as suggested by TNT. Traditional subgroups are identified by color branches.

- **Stage-1 dataset with root cognacy** (*wind = veter*, *agni = ignis*): **(a)** proper IE [6 optimal trees with score 951 obtained], **(b)** IE-Samoyed [4 optimal trees with score 1088 obtained].
- **Stage-2 dataset without derivational drift** (*wind ≠ veter*, *agni = ignis*): **(c)** proper IE [1 optimal tree with score 977 and 8 suboptimal trees with score 978 obtained], **(d)** IE-Samoyed [1 optimal tree with score 1112 and 3 suboptimal trees with score 1113 obtained].
- **Stage-3 homoplasy-optimized dataset** (*wind ≠ veter*, *agni ≠ ignis*): **(e)** proper IE [5 optimal trees with score 980 obtained], **(f)** IE-Samoyed [10 optimal trees with score 1116 obtained].

For Stage-3, no topological discrepancies with and without Proto-Samoyed, i.e., between (e) & (f).
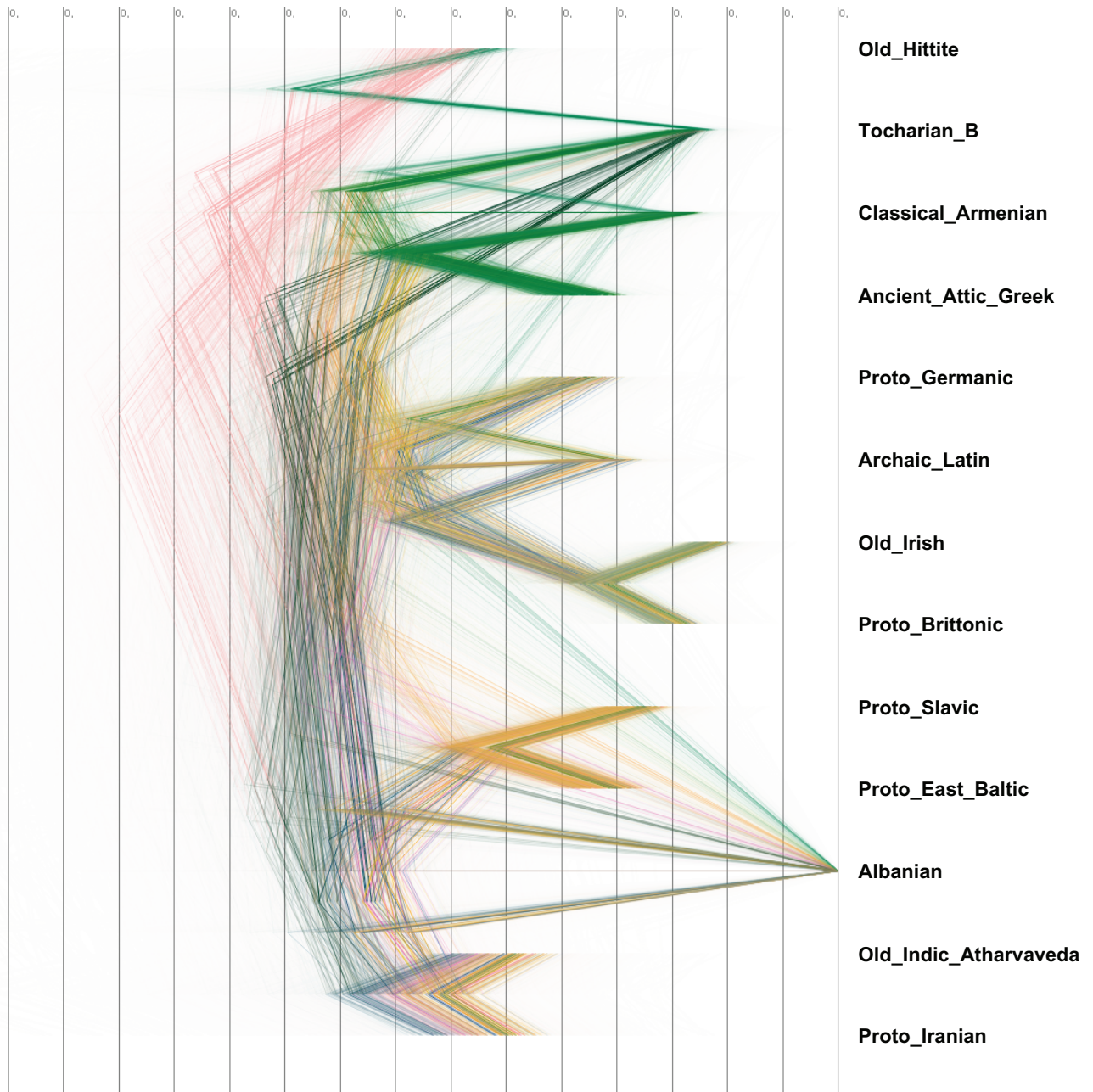
Fig. S4. DensiTree plots (Bouckaert and Heled 2014) for the consensus trees produced by the Bayesian MCMC analysis. **Stage-1 dataset with root cognacy** (*wind = veter, agni = ignis*): proper IE without Samoyed (see Fig. S2a for the 50% majority rule tree).
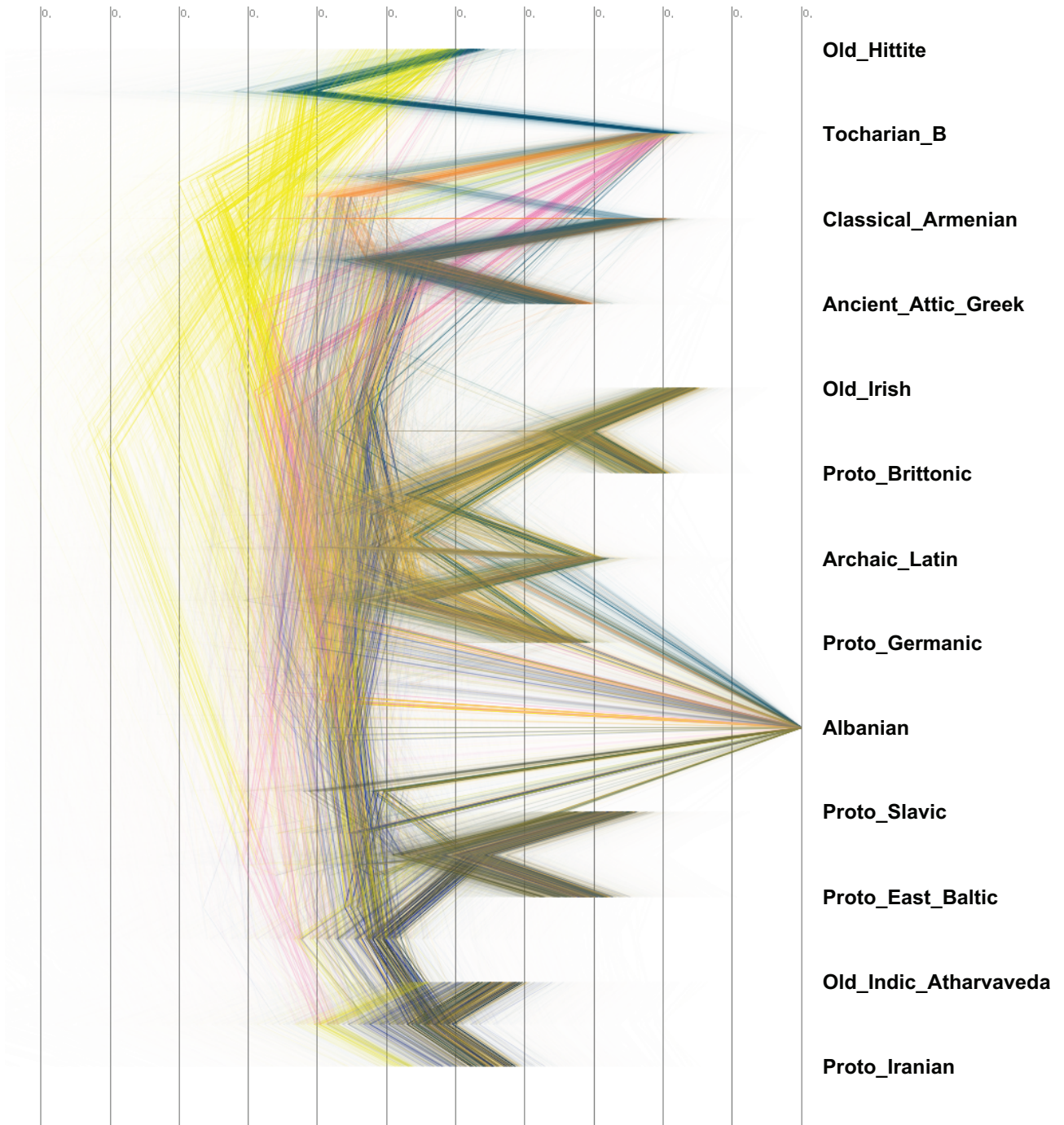
Fig. S5. DensiTree plots for the consensus trees produced by the Bayesian MCMC analysis. **Stage-2 dataset without derivational drift** (*wind ≠ veter, agni = ignis*): proper IE without Samoyed (see Fig. S2c for the 50% majority rule tree).

Fig. S6. DensiTree plots for the consensus trees produced by the Bayesian MCMC analysis. **Stage-3 homoplasy-optimized dataset** (*wind ≠ veter, agni ≠ ignis*): proper IE without Samoyed (see Fig. S2e for the 50% majority rule tree).
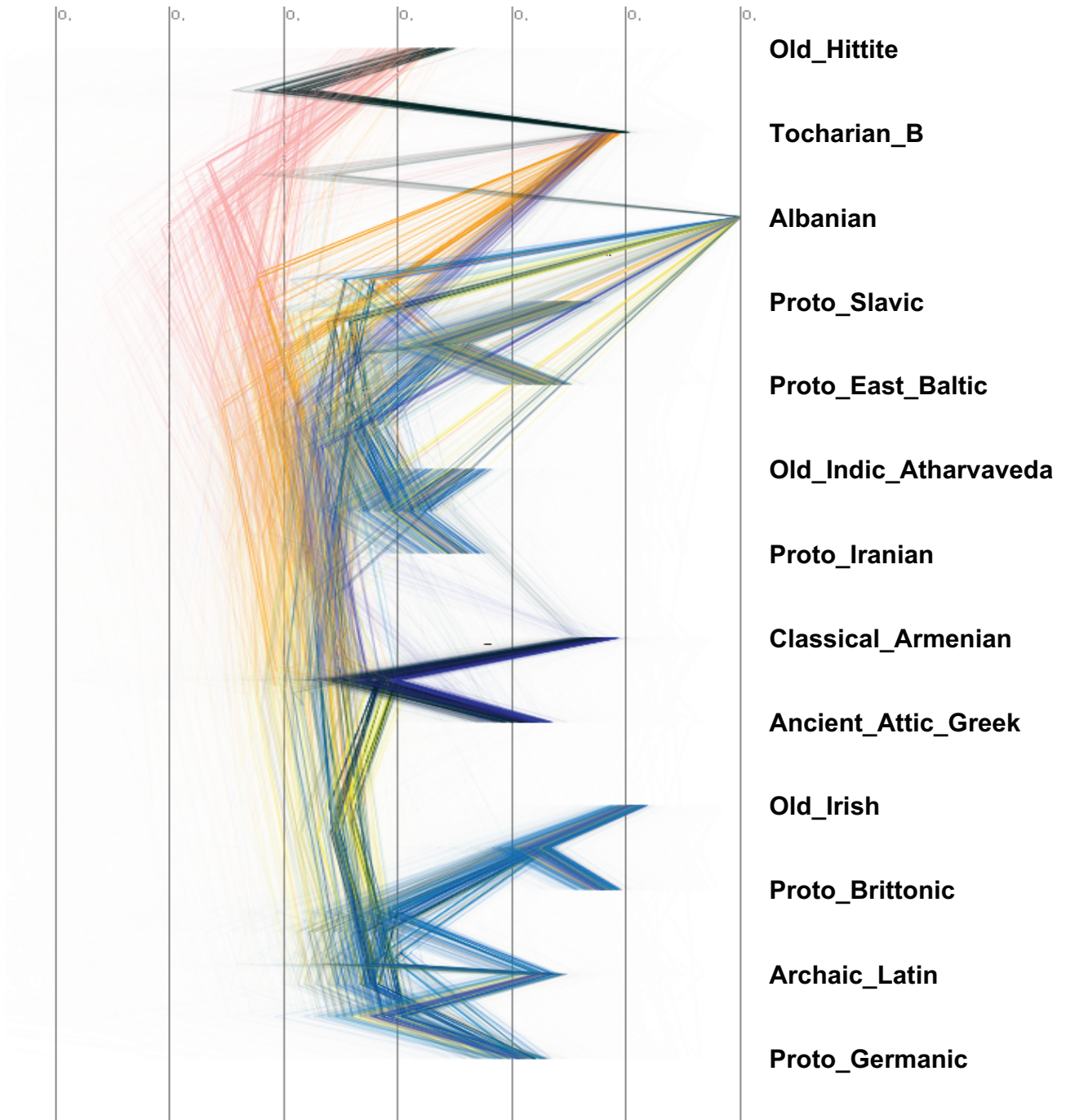
Bipartitions and their posterior probabilities for the Bayesian analysis

Tab. S1 represents bipartitions (splits) obtained by the Bayesian analysis for the Stage-3 dataset without Samoyed (MrBayes file *.parts; by default splits with posterior probability < 0.05 are not taken into account). The upper (green) part of the table contains the splits with posterior probability ≥ 0.5 which are summarized as the Indo-European tree Fig. S2e. The lower (yellow) part shows splits with posterior probability < 0.5, these splits are not present in the tree Fig. S2e, but are interesting in regard to alternative branching of Indo-European.

Tab. S1. Bipartitions (splits) and their posterior probabilities for the Bayesian analysis of the Stage-3 dataset without Samoyed (Fig. S2e): highlighted taxa in a row form a distinct clade.

ID	Hitt	Toch	Grk	Arm	Alb	Lat	Iri	Brit	Germ	Slav	Balt	Ind	Iran	Prob
14	.	.	.	.	.	.	*	*	.	.	.	.	.	1.00
15	.	.	.	.	.	.	.	.	.	.	.	*	*	1.00
16	.	.	.	.	.	.	.	.	.	*	*	.	.	1.00
17	.	.	.	.	.	*	*	*	*	.	.	.	.	0.99
18	.	.	*	*	.	.	.	.	.	.	.	.	.	0.94
19	.	.	*	*	*	*	*	*	*	*	*	*	*	0.80
20	.	.	.	.	.	.	.	.	.	*	*	*	*	0.76
21	.	*	*	*	*	*	*	*	*	*	*	*	*	0.70
22	.	.	.	.	*	.	.	.	.	*	*	*	*	0.54
23	.	.	.	.	.	*	.	.	*	.	.	.	.	0.44
24	.	.	*	*	.	*	*	*	*	.	.	.	.	0.43
25	.	.	.	.	.	*	*	*	.	.	.	.	.	0.35
26	*	*	.	.	.	.	.	.	.	.	.	.	.	0.25
27	.	.	.	.	*	*	*	*	*	*	*	*	*	0.24
28	.	.	.	.	.	.	*	*	*	.	.	.	.	0.21
29	.	.	.	.	*	.	.	.	.	*	*	.	.	0.18
30	.	.	*	*	.	*	*	*	*	*	*	*	*	0.15
31	.	*	*	*	.	.	.	.	.	.	.	.	.	0.15
32	.	.	.	.	.	*	*	*	*	*	*	*	*	0.11

Basic information on the language groups

Hittite (Alexei S. Kassian)

The Hittite list is adapted from Kassian 2011a. It is based on Old Hittite data, although it also includes a small number of terms ('feather', 'hair', 'far', 'heavy') that are attested only in Middle or New Hittite sources. The list is approximately dated back to 1650–1500 BC in accordance with the so-called middle chronology. There are some roots for which one can postulate the correspondence Hittite *s* : Luwian *t* : Nuclear IE Ø. In our list, it is observed in the items 'fingernail' and 'eye', but the actual instances are more numerous comprising such words from basic vocabulary as 'urine', 'tear(s)', 'oil', probably 'to live'. Cognation between at least the Hittite words for 'fingernail' & 'eye' and the corresponding Nuclear IE terms is currently widely accepted among Indo-Europeanists (see Tischler 2004: 737, 838; Kümmel 2008: 30–31; Kloekhorst 2008: 705, 723 for overview). However, various authors propose different explanations of the correspondence *s* : *t* : Ø. E.g., Ivanov (2001: 133) and independently Kassian and Yakubovich (2013: 22) introduce a specific Proto-Indo-European interdental phoneme *θ, while other authors suppose either a specific development of an initial "laryngeal" (*h_3-) or the *s*-mobile.

Tocharian B (Anna Kuritsyna)

The Tocharian B list is adapted from Kuritsyna 2017. The corpus for Tocharian B is dated back to the 5th–11th centuries AD. Since Tocharian A is generally believed to have no longer been in active use by the time it was written down, we have decided to represent the Tocharian branch primarily by evidence from Tocharian B.

Ancient Attic Greek (Alexei S. Kassian)

Ancient Attic Greek is formally very close to a protolanguage of modern Greek dialects excepting modern Tsakonian. The list is based on the idiolect of Plato (ca. 425–348/347 BC) as it is compiled in Kassian 2011b. Three items — 'bark', 'fat', 'louse', 'worm' — are missing from Plato, in these cases it does not seem particularly risky to fill the slot with the corresponding term from other Attic authors of the same period: Xenophon (ca. 430–354 BC), Sophocles (497/496–406/405 BC), Aristophanes (ca. 446–386 BC). Nevertheless the slot

'worm' remains empty due to scantiness of data. The only case within the 110-item word-list where Plato demonstrates an innovative lexeme as compared to Modern Greek is Plato *kérk-o-s* ⟨κέρκος⟩ 'tail' vs. Pan-Greek ⟨οὐρή⟩.

Classical Armenian (Petr A. Kocharov)

The Classical Armenian list is adapted from Kocharov 2017. The main part of our corpus is dated back to the 5th century AD.

Albanian (George Starostin)

The Albanian list is adapted from G. Starostin 2011 with additions and corrections concerning verbal suppletion and loanword detection. The list is based on the Literary Albanian language, i.e., on the Tosk dialect. Comparison with the Gegh dialects demonstrates that lexicostatistical differences between Tosk and Gegh are minimal (if at all existent), which makes reconstruction of a Proto-Albanian wordlist unnecessary.

Archaic Latin (Mikhail N. Saenko)

The Archaic Latin list is adapted from Saenko 2015. It is based on the plays of Titus Maccius Plautus (*ca*. 250–184 BC). Formally Plautus' Latin is very close to a proto-language for modern Romance languages.

Celtic (Old Irish and Proto-Brittonic) (Mikhail Zhivlov)

Due to fragmentary nature of Continental Celtic evidence, practical reconstruction of a Proto-Celtic list is impossible. We also decided not to reconstruct a Proto-Insular Celtic list because of the binary structure of Insular Celtic (Goidelic vs. Brittonic) and numerous cases where both branches have different innovations with respect to PIE.

The Old Irish list is based on contemporary Old Irish sources, the largest ones being Würzburg Glosses (*ca*. 750 AD), Milan Glosses (*ca*. 800 AD) and St. Gall Glosses (*ca*. 850 AD). In the rare cases where the word is not found in Old Irish proper, or its meaning is not clear from attestations, we are forced to fall back on data from later stages of the lan-

guage. The Old Irish Swadesh list in Lucht 2007 proved very useful, although on a number of points our choice differs from that of Lucht.

The Proto-Brittonic list is reconstructed based on data from various lexicographic sources on modern and medieval Brittonic languages. The phonology of Proto-Brittonic forms largely follows that of Schrijver's (1995) Late Proto-British.

Proto-Germanic (George Starostin)

Onomasiological reconstruction is based on the data collected in G. Starostin 2016 as well as on various lexicographic sources on the Germanic languages.

Proto-Slavic (Alexei S. Kassian)

Onomasiological reconstruction is normally based on the data collected by Saenko (2013) and Kassian & Anna Dybo (see linguistic supplement in Kushniarevich et al. 2015). We proceed from the tree published in Kushniarevich et al. 2015 which implies an initial three-way split into the West, East and South subgroups. We adhere to the modern reconstruction of the Proto-Slavic vowels as it is described in Kiparsky 1963; Kassian 2001 and some other authors. In brackets, the traditional Slavistic reconstruction is offered. The only exception is the Proto-Slavic monophthong *i: and the diphthong *ey — both are transcribed as *i: (= traditional Slavistic notation ⟨*i⟩), because the currently available inner Slavic evidence for discriminating between Proto-Slavic *i: and Proto-Slavic *ey is relatively weak and requires additional investigation.

Proto-East Baltic (Alexei S. Kassian)

Out of the Baltic lects, we are forced to confine ourselves to two living East Baltic languages — Lithuanian and Latvian — which are well documented and thus suitable for lexicostatistical analysis. The languages of the West Baltic subgroup — Old Prussian and some others — are so poorly documented that they can only serve as a source of external *comparanda*, which help to make a choice when there are lexical discrepancies between Lithuanian and Latvian. Onomasiological reconstruction is normally based on data collected by Kassian and Anna Dybo (see linguistic supplement in Kushniarevich et al. 2015)

as well as on various lexicographic sources on the Baltic languages. We ignore accentual features in our reconstruction.

Vedic (Atharvaveda) (Artem A. Trofimov)

The Vedic list is adapted from Trofimov 2016. It is based on Atharvaveda-Samhita. It must be noted that the Rigveda is older but has more lacunae within the 110-item wordlist. Formally Atharvaveda is very close to a protolanguage of the modern Indo-Aryan languages. The only word missing from Atharvaveda as well as from Rigveda and thus taken from a later source is, surprisingly, 'thin'.

Proto-Iranian (Artem A. Trofimov)

This is the most problematic portion of the dataset due to our insufficient knowledge of the phylogeny of the Iranian group. Recent research has shown that the traditional division of Iranian languages into West and East Iranian is outdated and does not reflect historical reality (N. Sims-Williams 1996; Windfuhr 2009; Korn 2016; Korn 2019). First, the Eastern Iranian group certainly does not represent genetic unity. Sims-Williams supposes that

> it does not seem possible to regard the Eastern Iranian group as a whole — even disregarding Parachi and Ormuri — as a genetic grouping … if one reconstructs "proto-eastern-Iranian" in such a way as to account for all the features of the group, it proves to be identical to the "common Iranian" reconstructible as the ancestor of the whole Iranian family. (N. Sims-Williams 1996, 651)

Moreover, A. Korn shows that the division into North-Western and South-Western Iranian based on mainly phonological differences between Middle Persian and Parthian as described by Tedesco (1921), is also problematic if one considers New Iranian languages thought to be descendants of these two Middle Iranian idioms, so there are no ancestral nodes for NW and SW Iranian:

> most New WIr. languages do not arrange themselves on one side of the division, sharing instead some features with Parthian, but others with (Middle) Persian. (Korn 2016, 405)

Further, Korn demonstrated that for the Old Iranian period, the distinction between marginal languages and the central ones existed; she tentatively reconstructs Central Iranian comparing shared innovations of Bactrian and Parthian and thinks that this subgroup can include some New Iranian languages (Zazaki) (Korn 2016, 419–431).

Nevertheless, recognizing the problem we use the terminology of traditional classification, because there is no alternative consensus classification and up-to-date tree of the Iranian clade (working on the Proto-Iranian Swadesh list we extensively used the data from Rastorgueva and Edelman 2000–, the authors of this compendium also adhere to the traditional Iranian classification). Traditional classification with two primary branches — Eastern Iranian languages and Western Iranian languages — together with the external *comparanda* from the closely related Indian group are sufficient for onomasiological reconstruction of the majority of Proto-Iranian Swadesh items. Our study is based on the data collected in Trofimov and Belyaev 2014 as well as on various lexicographic sources on the Iranian languages.

Proto-Samoyed (Mikhail Zhivlov)

Representing a group distantly related to Indo-European, this list is used as an outlier in the phylogenetic analysis. Samoyed is the most divergent branch of the Uralic language family. There is no consensus on the internal classification of Samoyed languages. We provisionally accept E. Helimski's classification, according to which Northern Samoyed languages — Tundra and Forest Nenets, extinct Old Eastern Yurak (sometimes erroneously called Yurats), Tundra and Forest Enets and Nganasan — form a subgroup, while the so called Southern Samoyed languages (Mator, Kamass and Selkup) represent three independent branches of the Samoyed group (Helimski 1982). Nevertheless, any lexical isogloss unique to "Southern" languages can in principle result from a later areal development, so that attestation in at least one Northern and one "Southern" language is crucial for assuring that a word was present in Proto-Samoyed. Of course, a word with good etymological parallels in other branches of Uralic must be reconstructed for Proto-Samoyed even if it is attested only in one Samoyed language. Our Proto-Samoyed reconstruction follows Janhunen 1977 with important additions and corrections by Helimski (1997: 68–70; 2005). However, we do not distinguish in our transcription between front and back reduced vowels, since this distinction is morphophonemic rather than phonological.

Dating of the nodes and chronological constraints

Tab. S2. Dating of the nodes and chronological constraints

Taxa	Constraints for Bayesian analysis	Strict dates for StarlingNJ analysis	Comments
Old Hittite	1650–1500 BC	1550 BC	Traditionally the Old Hittite language is associated with the Old Hittite kingdom, which existed from the king Hattusili I to the king Telipinu covering the period *ca*. 1650–1500 BC according to the so-called middle chronology (Watkins 2008: 6; Yakar 2012: 75, 78).
Tocharian B	400–900 AD	650 AD	The period *ca*. 5th–9th c. AD is a commonly accepted estimate for the Tocharian B corpus (Peyrot 2013: 1; Adams 2017: 452).
Ancient Attic Greek	375 BC	375 BC	The Attic wordlist is based on Plato's (428/7–348/7 BC) works.
Classical Armenian	400–500 AD	450 AD	Our wordlist is based on the Classical Armenian text mostly compiled in the 5th c. AD, first of all on the Armenian translation of the Bible (Clackson 2008: 125).
Albanian	1950 AD	1950 AD	Modern literary Albanian (based on the Tosk dialect) is used.
Archaic Latin	200 BC	200 BC	The Latin wordlist is based on Plautus' (ca. 250–184 BC) works.
Old Irish	700–900 AD	800 AD	Our wordlist is based mainly on contemporary Old Irish sources, the largest ones being Würzburg Glosses (*ca*. 750 AD), Milan Glosses (*ca*. 800 AD) and St. Gall Glosses (*ca*. 850 AD).
Proto-Brittonic	300–600 AD	450 AD	According to the commonly accepted view, "[t]he Brythonic dialects began to diverge into West British (> Welsh) and South-West British (> Cornish and Breton) in about the fifth century AD, but probably remained mutually intelligible for several centuries" (P. Sims-Williams 2017: 352).

Taxa	Constraints for Bayesian analysis	Strict dates for StarlingNJ analysis	Comments
			The fifth century date is consistent with the historical evidence, since the Anglo-Saxon invasion of Britain in the 5th c. AD is expected to have been a natural trigger for the divergence of Proto-Brittonic. Our lexicostatistical estimation of the divergence between West Brittonic and South-West Brittonic places it somewhat earlier, at c. 300 AD.
Proto-Germanic	500–300 BC	400 BC	Proto-Germanic, before its split, can be dated back to 500–300 BC (Ringe 2017: 241) or more precisely 500 BC as proposed in Bousquette and Salmons 2017: 389. The common view is that this language is to be associated with the Jastorf archaeological culture of 600–1 BC (Mallory and Adams 1997: 222–223).
Proto-Slavic	1–300 AD	100 AD	The lexicostatistical estimation of the initial three-way split of Proto-Slavic is 100 AD as obtained in Kushniarevich et al. 2015; later, ca. 500 AD the three Slavic branches started to divide into more narrow clusters. This is in accordance with the common view that the Slavs occupied a relatively large territory in the first half of the 1st millennium AD (Sussex and Cubberley 2006: 19) and it is likely that the Slavic population was no longer homogenous at that stage (Sedov 1995: 5), further see linguistic supplement, p.18 in Kushniarevich et al. 2015. Thus the period 1–300 AD is a reasonable approximation for the divergence of Proto-Slavic.
Proto-East Baltic	400–1 BC	200 BC	The lexicostatistical estimation of the divergence between the Lithuanian and Latvian branches of the East Baltic clade is 200 BC (Kushniarevich et al. 2015). Among experts, this issue is very rarely discussed, so there is no consensus view on the date of the Lithuanian-Latvian split as well as on archaeological cultures that could be associated with the Lithuanian-Latvian protolanguage. As concluded in Adams and Mallory 1997a: 50, it is likely that the split of Proto-

Taxa	Constraints for Bayesian analysis	Strict dates for StarlingNJ analysis	Comments
			Baltic into the East and West branches can be dated back to ca. 1000 BC. So the assumed subsequent divergence of Proto-East Baltic in 200 BC does not contradict the available data.
Old Indic (Atharvaveda)	1200–1000 BC	1100 BC	Atharvaveda is traditionally dated back to 1200–1000 BC (Witzel 2003: 68; Kulikov 2017: 215).
Proto-Iranian	1500–1000 BC	1300 BC	This is the most problematic and ambiguous case in respect of absolute dating. It seems that 1500–1000 BC is a reasonable estimate for the initial breakup of Proto-Iranian, cf. Adams and Mallory 1997b: 311; N. Sims-Williams 2017: 264.
Proto-Samoyed	950–750 BC	800 BC	Our lexicostatistical estimation of the divergence of Proto-Samoyed into daughter languages is 800 BC. This date coincides with the lexicostatistical date of the split of Proto-Ob-Ugric into Mansi and Khanty branches, indicating some ethnohistorical processes in Western Siberia at that time. Identification of Proto-Samoyed community with one of the Western Siberian archaeological cultures remains a moot question, one of the principal candidates being the Kulay culture (Napolskikh 1997: 83–84) which is traditionally dated back to the end of 6th century BC — 5th century AD (Chindina 2001). Since this identification remains controversial, we are forced to use the lexicostatistical date here.

Additionally, for the Bayesian MCMC analysis, we have to predefine the root age, i.e., the temporal range in which the first bifurcation occurred.

For the proper IE dataset, we set the hard limits as 3500 BC and 8500 BC. This means that we say to the program that the first split within the IE family must fall within the range of 8500–3500 BC. The upper limit 3500 BC follows from the fact that wheel terminology is not reconstructible for the Indo-Hittite level; the most archaic term seems to be *$k^w ek^w lo$-

'wheel' shared by Tocharian and Inner IE, but not by Anatolian. It is thus likely that the Anatolian branch split off before the invention and quick spread of wheeled transport in 4000–3500 BC (see most recently Anthony and Ringe 2015). The lower limit 8500 BC is based on the term for domesticated cattle which is reconstructible for Indo-Hittite including Anatolian and Tocharian (*$g^w ow$- 'cow'). The early traces of wild aurochs domestication found in the Near East date back to *ca.* 8500 BC (Helmer et al. 2005; Hongo et al. 2009).

For the IE-Samoyed dataset, the range is set as 8000 BC (minimal) and 18,000 BC (mean), meaning that the split into the IE and Uralic branches must have occurred around 8000 BC or earlier, but not likely earlier than 18,000 BC, although it is not excluded. Both dates are inevitably arbitrary, but we consider them reasonable.

Linguistic comments on individual Swadesh forms

Note on transcription. All linguistic data in the present document are encoded in the unified transcription system of the Global Lexicostatistical Database project, which is generally based on the IPA alphabet, with a few specific discrepancies, e.g., *c* stands for IPA *ts*, *š* for IPA ∫ (http://starling.rinet.ru/new100/UTS.htm). Traditional or orthographic representations are enclosed in ⟨angle brackets⟩.

1 'all'.

Old Irish. *ulʸe* ⟨uile⟩ means both 'all = omnis' and 'all = totus'. Goes back to **ol-yo-*, possibly cognate with Germanic **alla-*.

Proto-Brittonic. **hol:* is retained in all daughter languages. Goes back to **sol-no-*, possibly cognate with Oscan *sullus* 'omnes', and, with another suffix, Old Indic *sˈarw-a-* ⟨sárva-⟩.

Proto-Germanic. **all-a-* is retained in all three subgroups.

Proto-Slavic. **vix-* ⟨**vьsь, *vьxъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **vis-a-* is retained in Lithuanian and Latvian.

Proto-Iranian. **wic-wa-* is the best candidate: its descendants mean 'all' in Avestan and in some Middle Iranian languages (like Sogdian). This stem has Indo-Aryan cognates with the same meaning. In many Iranian languages (both western and eastern) **wic-wa-* was superseded by the stem **ham-a-* with the original meaning 'the same'; in Parachi and Ormuri it has been replaced by reflexes of **har-wa-* 'whole (*totus*)'.

Proto-Samoyed. **tük-* (Janhunen 1977: 168) is attested in Northern Samoyed and Mator.

Etymological notes. Similarity between Proto-Balto-Slavic **vis-a-* 'all' and Proto-Indo-Iranian **wic-wa-* 'all' can be accidental: since 'all' is a highly unstable item in the Indo-European family, an abundance of various forms increases the probability of chance phonetic similarity. On the other hand, Balto-Slavic **vis-a-* 'all' and Indo-Iranian **wic-wa-* can be treated as etymological cognates containing the common virtual root **wikʸ-*, if one as-

sumes the occasional suffixal pattern *wik^y-so-* for Balto-Slavic, correspondingly the Indo-Iranian form would originate from *wik^y-wo-*. We prefer to accept the first solution.

2 'ashes'.

Old Irish. *l:ua̯θ^y* ⟨*lúaith*⟩ means both 'ashes' and 'dust'.

Proto-Brittonic. **l:ü:du* is retained in all daughter languages.

Proto-Germanic. **ask-o:n* is retained in all three subgroups.

Proto-Slavic. **pepel-u ~ *papel-u* ⟨**pepelъ ~ *popelъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **pel-en-a-* is retained in Lithuanian and Latvian.

Proto-Iranian. **a:tr-ya-* is retained in Avestan and in many other Eastern Iranian languages. It is derived from **a:tar-* 'fire'.

Proto-Samoyed. **kimä* (Janhunen 1977: 70) is attested in Northern Samoyed and Selkup.

Etymological notes. Slavic **pepel-u* and Baltic **pel-en-a-* represent different substantive formations from the verbal root **polə-* (**polh₁-*) 'to flame up, singe (*vel sim.*)'. Because of this we mark the Slavic and Baltic forms with separate indexes in the derivational drift-free dataset (Stage-2).

3 'bark'.

Old Irish. *r:u:sk* ⟨*rúsc*⟩.

Proto-Brittonic. **r:i:sk* is retained in Welsh. Breton ⟨*rusk*⟩ 'bark' and Cornish ⟨*rusc*⟩ 'id.' have deviant vowel reflexes, making one suspect that these words are borrowings from Goidelic.

Proto-Germanic. Not attested in Gothic. Proto-Germanic **bark-u-* functions as the main equivalent for 'bark' in Scandinavian languages (from where it has also been borrowed into English) and is occasionally found in Frisian and Low German dialects as well. However, the most common West Germanic root is **rend-o:n* (Old English *rind*, Old High German

rinta, etc.). Both roots may be formally included into comparison, although both have possible internal Germanic / Indo-European etymologies that define them as innovations.

Proto-Slavic. **kar-aː ~ *skar-aː ⟨*kora ~ *skora⟩* is retained in all three subgroups.

Proto-East Baltic. **žeyv-eː* is retained in Lithuanian and the Latvian dialects. Probably a deverbative with the *v*-suffix, although the starting root **žey-* is not attested elsewhere in Baltic.

Proto-Iranian. **pawasta-* is likely to be analyzed as the past participle **pa=was-ta-* from the verbal root **was-* 'to wear clothes, be dressed' with prefix **(a)pa-*. The first meaning of this form is 'skin'; polysemy 'skin / bark' is found in many Iranian languages and can be reconstructed for the proto-language. The stem **pawasta-* is retained in Avestan with the meaning 'skin (especially on the head of a human)' and, with polysemy, in Southwestern (Middle Persian ⟨*pōst*⟩ 'skin / hide', Modern Persian ⟨*pust*⟩ 'skin / hide / bark / peel / (nut)shell') and several Eastern Iranian languages (Wakhi ⟨*pist*⟩ 'skin / hide / bark', Sanglechi ⟨*pask*⟩ 'id').

Proto-Samoyed. **kasʋ* (Janhunen 1977: 65) is attested in Northern Samoyed, Kamass and Selkup.

4 'belly'.

Old Irish. *bruː ⟨brú⟩*. Polysemy: 'abdomen, belly; bowels, entrails; womb'.

Proto-Brittonic. **torː* is retained in Cornish. Cognate to Welsh ⟨*tor*⟩ '(upper part of the) belly' and Old Breton ⟨*tar*⟩, gl. 'uentrem'. The main Welsh word for 'belly' is ⟨*bol*⟩, whose original meaning is 'bag, pouch'; Breton has ⟨*kof*⟩ of unclear etymology.

Proto-Germanic. A highly unstable item. At least three candidates may be suggested, none of them, however, precisely matching the required semantics of 'abdomen': (a) **wamb-oː*, well attested in Gothic and West Germanic in application to either 'womb' or 'abdomen' in general; in West Scandinavian dialects, it seems to be a "vulgar" term for 'belly'; (b) **kwiθu-*, probably meaning 'stomach' in Gothic, 'belly / womb' in Scandinavian, and 'womb' in old West Germanic languages; (c) **mag-oːn*, usually with the specific semantics of 'stomach' in old languages, but extending to 'abdomen' as well in some modern languages (e.g., Danish). It may be assumed that Proto-Germanic did not have a single lexical

equivalent for the notion of 'belly' as 'body segment', and while there is no solid evidence for item (c) ever fulfilling that function in the oldest layers of Germanic, both (a) and (b) are probably eligible for inclusion as synonyms.

Proto-Slavic. An unstable item. The best candidate is *červ-a* ⟨*červ-o*⟩ which means 'belly' in the South (Old Church Slavonic) and East (Old Russian, Ukrainian dialects) subgroups. Another potential candidate is *bryoːx-a* ⟨*brjuxo*⟩ used in modern West lects and in some Russian dialects. *červ-a* seems preferable from the chronological point of view, whereas *bryoːx-a* could be a later parallel development among the West and East lects. Nevertheless, we treat both as technical synonyms.

Proto-East Baltic. *veːder-a-* means 'belly' in Latvian and 'entrails' in Lithuanian; its antiquity is proven by the semantics 'belly' in Old Prussian.

Proto-Iranian. The stem *udar-a-* is retained in Avestan, Sogdian and in many Eastern Iranian languages. The choice is additionally supported by the Indo-Aryan cognates.

Proto-Samoyed. It is hard to choose between *nançŏ* (Janhunen 1977: 20) and *pärkä* (Janhunen 1977: 122). *nançŏ* is retained in Northern Samoyed, Mator, Kamass and Selkup, *pärkä* — in Mator and Selkup (with derivatives in Northern Samoyed). The semantic difference between the two words in Mator and Selkup is difficult to establish.

Etymological notes. There are two Inner IE terms for 'belly' in criss-crossed configuration. (1) The root cognacy between Old Irish *bruː* ⟨*brú*⟩ (< Celtic *brus-on-* < IE *bʰrus-on-*) and Slavic *bryoːx-a* ⟨*brjuxo*⟩ (< IE *bʰrews-o-*). (2) Latin *wˈenter* ⟨*venter*⟩, Baltic *veːder-a-*, Indo-Iranian *udar-a-*. The second one has the clear advantage in terms of distribution and morphological uniformity. On the contrary *bruː* / *bryoːx-a* looks like a typical derivational drift, although we have no formal criteria to determine it. So we are forced to assign the same index to *bruː* and *bryoːx-a* in the derivational drift-free dataset (Stage-2), but mark these forms with two different indexes in the homoplasy-optimized dataset (Stage-3).

5 'big'.

Old Irish. *moː-r* ⟨*mór*⟩.

Proto-Brittonic. *mɔːr is retained as the main word for 'big' in Welsh, but has cognates in Cornish and Breton as well. The main word for 'big' in Cornish and Breton goes back to *bras, whose original meaning, preserved in Welsh, was 'thick, fat'.

Proto-Germanic. *mekil-a- 'big' is well attested in most of the ancient Germanic languages, although the word has been subsequently replaced in all modern languages.

Proto-Slavic. *vel-iːk- ⟨*velikъ⟩ is retained in all three subgroups.

Proto-East Baltic. *leyl-a- has the general meaning 'big' in Lithuanian dialects and Latvian; its relation to the phonetically similar stems with the meaning 'lean, skinny' is unclear. In Literary Lithuanian, 'big' is expressed by a suffixal derivative from *did-i- 'great'.

Proto-Iranian. The root *maʒ- and its derivatives *maʒ-ant- and *maʒ-a-na- are retained in Avestan, Sogdian, Southwestern (Middle Persian ⟨māzanīk⟩), Northwestern (Kurdish ⟨mazin⟩, Balochi ⟨mazan⟩) and in some Eastern Iranian languages (for example, Manichean Sogdian maz-eːχ ⟨mzyγ, mzyx, mzyyx⟩ < comparative degree *maʒ-iya-h-). It has an undoubted Indo-European etymology and exact matches in Old Indic.

Proto-Samoyed. *ʋr-kʋ (Janhunen 1977: 19) is retained in Northern Samoyed, Mator, Kamass and Selkup. The word is derived from *ʋrə̆ 'magnitude'.

6 'bird'.

Old Irish. eːn ⟨én⟩.

Proto-Brittonic. *edn is retained in Welsh, Cornish and East Breton dialects. Welsh also has ⟨aderyn⟩ 'bird', singulative from ⟨adar⟩ 'birds'. This word goes back to the same IE root *pet- as *edn.

Proto-Germanic. *fugl-a- is retained in most ancient and modern languages.

Proto-Slavic. The root *put- ⟨*pъt-⟩, modified with various suffixes, is retained in all three subgroups.

Proto-East Baltic. The root *put- modified with various suffixes means 'bird' in Latvian and 'chicken' in Lithuanian. Cognate to the Slavic root for 'bird'.

Proto-Iranian. The stem *wi- is retained in Avestan and Middle Persian and is supported by Vedic evidence. In many Iranian languages, 'bird' is expressed with help of *mr̥ga- which looks like a later and sometimes contact-driven innovation for this meaning. Apparently the Proto-Iranian meaning of *mr̥ga- was 'a big bird' as attested in Avestan mr̥ga ⟨marəga⟩ 'a big bird' (like eagle, han). Indo-Aryan cognates have the semantics 'wild animal'.

Proto-Samoyed. *sʋrmʋ (Janhunen 1977: 136) is a general word for 'bird' and '(wild) animal'. No specific word for 'bird' can be reconstructed.

7 'to bite'.

Old Irish. There is no verb 'to bite' in Old Irish. The meaning is expressed by various periphrastic constructions probably of innovative nature.

Proto-Brittonic. Not reconstructible. Each language has its own word for this meaning, and external comparison does not help in this case.

Proto-Germanic. *biːt-an- is retained in most ancient and modern languages.

Proto-Slavic. *kans- ⟨*kǫs-⟩ is retained in all three subgroups.

Proto-East Baltic. *kand- is retained in both Lithuanian and Latvian.

Proto-Iranian. An unstable concept. The root *danc- seems the only reliable candidate. Although it is not attested in finite forms, it was retained in the Khotanese Saka participle ⟨duṣṭa⟩ 'stung' and some nominal derivatives such as Avestan ⟨tiži.dąsura-⟩ 'biting sharply (of dog)'. The main advantage of *danc- is that it is supported by exact Indo-Aryan cognates.

Proto-Samoyed. *saç- (Janhunen 1977: 136–137) is retained in Northern Samoyed and Selkup.

8 'black'.

Old Irish. duv ⟨dub⟩.

Proto-Brittonic. *duv is retained in all languages.

Proto-Germanic. *svart-a-* is retained in most ancient and modern languages.

Proto-Slavic. *čir-n-* ⟨*čьrnъ⟩ is retained in all three subgroups.

Proto-East Baltic. External comparison strongly suggests that the original Proto-Baltic stem was **kirs-n-a-* 'black' which is only retained in Old Prussian. In both Lithuanian and Latvian, innovative items are used: (1) **yo:d-a-* meaning 'black' in Lithuanian, 'evil spirit, demon' in Latvian, (2) **me:l-n-a-* which means 'black' in Latvian, being derived from Proto-Baltic *me:l-a-* '(dark) blue'. The first option is clearly preferable.

Proto-Iranian. **cya:-wa-* is retained in Avestan and the majority of Middle and Modern Iranian languages.

Proto-Samoyed. Not reconstructible: none of the roots used in this meaning in various Samoyed languages can be safely projected on the Proto-Samoyed level.

9 'blood'.

Old Irish. *fulʸ* ⟨fuil⟩ is the main unmarked word for 'blood'. Distinct from *kru:* ⟨crú⟩ 'gore, blood', used mainly of blood shed in a battle.

Proto-Brittonic. **gwɛ:d* is retained in all daughter languages.

Proto-Germanic. **blo:d-a-* is retained in all ancient and modern languages.

Proto-Slavic. **kru:*, obl. **kruv-* ⟨*kry, *krъv-⟩ is retained in all three subgroups.

Proto-East Baltic. A non-trivial case with two competing roots. First, the suffixed stem **kraw-ya-* ~ **kraw-ya:* which means 'blood' in Lithuanian and Old Prussian respectively. This stem looks like a substantivized adjective from the root cognate to Proto-Slavic **kru:* 'blood' (with another ablaut grade). Second, **asin-i-* ~ **asen-i-* meaning 'blood' in Latvian, not attested in other lects; it is very likely that **asen-* continues the main Indo-European root for 'blood'. Since the external evidence for the antiquity of **asen-i-* is solid and, on the other hand, **kraw-ya-* does not represent a direct morphological match to Slavic **kru:* 'blood', it is most likely that **kraw-ya-* is an areal Baltic innovation which spread either from Lithuanian to Old Prussian or *vice versa*.

Proto-Iranian. **wah-un-i-* ~ **wah-un-a-* is very stable in Iranian languages. It was probably derived from the PIE root **wes-* 'wet'.

Proto-Samoyed. *kɘm* (Janhunen 1977: 65) is retained in all Samoyed languages.

10 'bone'.

Old Irish. *knaːṽʸ* ⟨*cnáim*⟩.

Proto-Brittonic. *as-kurn,* retained in all daughter languages, is a compound whose first element continues PIE word for 'bone', while the second element is the old word for 'horn'.

Proto-Germanic. *bayn-a-* is retained in most ancient and modern languages (although the word 'bone' is not attested in Gothic).

Proto-Slavic. *kast-i* ⟨*kostь*⟩ is retained in all three subgroups.

Proto-East Baltic. *kawl-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *ast-* or its derivatives are retained in Avestan and the majority of middle and modern Iranian languages.

Proto-Samoyed. *lɘ* (Janhunen 1977: 82), retained in all Samoyed languages, goes back to Proto-Uralic *lɨwi* 'bone'.

11 'breast'.

Old Irish. It is hard to choose between two synonyms: *uxt* ⟨*ucht*⟩ 'breast, bosom' and *brunʸːe* ⟨*bruinne*⟩ 'breast, bosom, chest'. Both words are distinct from *kʸiːx* ⟨*cích*⟩ 'pap, (the female) breast'. ⟨*ucht*⟩ may be cognate with Latin ⟨*pectus*⟩, while ⟨*bruinne*⟩ is cognate with Brittonic *bronː* and Germanic *brewsta-*.

Proto-Brittonic. *bronː* is retained in Welsh and Cornish. Replaced in Breton by ⟨*bruched*⟩, borrowed from Old French.

Proto-Germanic. *brewst-a-*, retained in all modern languages, is equally applicable (at least in all ancient languages) to 'male breast (chest)' and 'female breast'.

Proto-Slavic. There are two candidates intertwined within the subgroups: (1) *grand-i* ⟨*grọdь*⟩, and (2) *pirs-i* ⟨*pьrsь*⟩. Slav. *grand-u ~ grand-aː* ⟨*grọdъ ~ grọda*⟩ 'lump, clod,

compact elevation' suggests the semantic development 'lump' > 'breast' for *grand-i ⟨*grǫdъ⟩, whereas ⟨*pьrsъ⟩ possesses external IE cognates with such meanings as 'rib' and so on, which prove that the anatomic semantics is primary for ⟨*pьrsъ⟩. On the other hand, the existence of Finnic *rinta 'breast, chest' theoretically can point to a virtual Baltic form *grind-a the Finnic term was borrowed from. However, even if such word existed in an extinct Baltic lects which influenced neighboring Proto-Finnic, the Slavic and Baltic stems demonstrate different ablaut grades, so it is most likely that we are dealing with parallel semantic developments in Slavic and Baltic.

Proto-East Baltic. *kru:t-i- or its suffixed derivatives are retained in both Lithuanian and Latvian.

Proto-Iranian. *war-ah-, found in Avestan and some other languages (for example, Balochi), has a parallel in Old Indian ʾur-as- ⟨úras-⟩. The original meaning of Proto-Iranian *fšta:n-(a)- is 'woman's breast / nipple'.

Proto-Samoyed. *sünsə̆ (Janhunen 1977: 144), retained in Northern Samoyed and Mator, has an external cognate in Hungarian and goes back to Proto-Uralic *sʸünsʸi.

Etymological notes. A criss-crossed configuration is observed in the Italic-Germanic-Celtic clade. Latin pʾektus ⟨pectus⟩ apparently corresponds to Old Irish uxt ⟨ucht⟩ (the same index in the derivational drift-free dataset, Stage-2). The second Old Irish synonym, brunʸ:e ⟨bruinne⟩ < Celtic *brunn-yo- < *bʰrus-n-yo-, is cognate to Proto-Brittonic *bron: < Celtic *brunn-a: < *bʰrus-n-a: and further to Proto-Germanic *brewst-a- < *bʰrews-t-o- with a different suffix pattern (the same index in the derivational drift-free dataset, Stage-2). The situation is not entirely clear. Provisionally, for the homoplasy-optimized dataset (Stage-3), we prefer to keep the match Latin pʾektus ⟨pectus⟩ : Old Irish uxt ⟨ucht⟩ as the most direct etymology and thus the main candidate for the status of the Proto-Italic-Germanic-Celtic term for 'breast'. In turn, Old Irish brunʸ:e ⟨bruinne⟩ (< *bʰrus-n-yo-) and Brittonic *bron: (< *bʰrus-n-a:) rather reflect a parallel development. Germanic *brewst-a- (< *bʰrews-t-o-) looks like a typical derivational drift, although we have no formal criteria to determine it. So we mark the Old Irish, Brittonic and Germanic forms with three different indexes in the homoplasy-optimized dataset (Stage-3).

12 'to burn (trans.)'.

Old Irish. *l:osk-əðʸ ⟨loscaid⟩*.

Proto-Brittonic. **l:osk-* is retained as the main verb with this meaning in Welsh and Cornish, and as a secondary synonym in Breton.

Proto-Germanic. **brenn-an-* is retained in all languages.

Proto-Slavic. According to the inner Slavic criteria, the main candidate is *pa:l-i:- ⟨*paliti⟩* which has the widest distribution as the basic verb for 'to burn (tr.)' in all three subgroups. It competes with **žeg- ⟨*žegti⟩* which is attested in some East, South and West (Polabian) languages. Despite its narrower distribution, however, **žeg-* has the advantage since, first, **žeg-* is attested in such ancient languages as Old Church Slavonic and Old Russian, second, **žeg-* has reliable external IE cognates with the same meaning 'to burn (tr.)'.

Proto-East Baltic. **deg-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **dag-* is retained in this or slightly altered meaning in Avestan, Middle Iranian and some Modern Iranian languages (for example, in Balochi ⟨*daχt*⟩ / ⟨*diž-*⟩ 'to brand', Wakhi perfective ⟨*pi=δn-*, *pi=dn-*, *pi=δiy̆n-*⟩ 'to catch fire, be ignited'); in modern Iranian languages its descendants often mean 'to hurt', 'to be ill'. Moreover, this stem has cognates in the Indo-Aryan group and other Indo-European languages. The second candidate — **sawk-* / **sauč-*, widespread in Iranian languages, has the original meaning 'to shine', also found in Avestan.

Proto-Samoyed. **kʋrə̆-* (Helimski 1997: 268–269) is attested only in Mator, but has an external etymology. According to Helimski, it can be compared to Proto-Uralic **karti-* 'to roast, to burn'. Other roots, used in this meaning in Samoyed languages, cannot be safely projected to the Proto-Samoyed level.

13. 'fingernail'.

Old Hittite. The etymological connection between *sankuw-ai-* and Inner IE **nogʰ-* / **n̥gʰ-* (**h₃nogʰ-* / **h₃n̥gʰ-*) 'fingernail' is widely accepted among modern Indo-Europeanists, although there is no consensus about phonological details of the reconstruction. In Kassian et al. 2015: 327, **mogʰ-* / **m̥gʰ-* is proposed as an IE proto-form.

Proto-Brittonic. **awi:n* is retained in all languages.

Proto-Germanic. **nagl-a-* is retained in all languages.

Proto-Slavic. **nag-ut-i* ⟨**nogъtь*⟩ is retained in all three subgroups.

Proto-East Baltic. **nag-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. Stems **naːx-u-na-* / **naːx-a-na-* and others, derived from the root **nax-* 'nail', are retained in all Iranian subgroups and have Indo-Aryan cognates. A very stable term in general. Only the oldest Iranian language, Avestan, has a word *sr-uː-* ⟨*srū-*⟩ / *sr-w-aː-* ⟨*sruuā-*⟩ with the polysemy 'horn / nail'. The elimination of the Proto-Iranian word for 'nail' in Avesta can be due to a very important role of nails in Zoroastrianism.

Proto-Samoyed. **kɞ̆tʋ* (Janhunen 1977: 55–56) is retained in all Samoyed languages.

14. 'cloud'.

Old Irish. *nːeːl* ⟨*nél*⟩. The word is sometimes viewed as a loan from Brittonic (cf. Welsh ⟨*niwl, nifwl*⟩ 'fog, mist, vapour, haze, cloud(s)'), because the expected Proto-Celtic **neblo-* would have given **⟨*nebul*⟩ in Old Irish. It is, however, possible that ⟨*nél*⟩ < **neglo-* is a result of contamination of expected **neblo-* with the reflex of PIE **mig^{h}la:* (**h₃mig^{h}leh₂*), otherwise not preserved in Celtic.

Proto-Brittonic. **kumul*, retained in all daughter languages, is a borrowing from Latin ⟨*cumulus*⟩ 'heap'.

Proto-Germanic. Difficult to reconstruct. From the semantic and distributional side, the best candidate is probably **melx-man* 'cloud', reflected in Gothic (*milx-ma*) and, in a different morphophonological variant, in Swedish *mol-(e)n* (< **mulx-na-*, according to Kroonen 2013: 363), although the Swedish equivalent is not the Common Scandinavian term for 'cloud'; both forms have Germanic and IE verbal parallels with the meaning 'to be overcast (of sky)'. The most widespread Scandinavian equivalent is Old Norse *skü:* and its descendants, possibly related to **ski:-n-an-* 'to shine' (Orel 2003: 341). West Germanic **wulk-an-* 'cloud' goes back to IE **welk-* 'moist'.

Proto-Slavic. **ab=valk-u* ⟨**ob-volkъ*⟩, literally 'enveloping', a prefixed substantive from the verb **velk-* 'to drag', is retained in West and South Slavic.

Proto-East Baltic. The best candidate is *debes-i-* meaning 'cloud' in Lithuanian and 'sky / cloud' in archaic Latvian. In Latvian, also the secondary stem *debes-ya-* 'cloud' is rarely used. In modern Latvian, *debes-i-* means only 'sky', whereas the meaning 'cloud' is expressed with a transparent new formation from the root *mak-* 'wet'.

Proto-Iranian. There are two Iranian stems, *ab-ra-* and *mayg-a-*, both attested in Avestan and all Iranian subgroups and having external cognates. It seems that the main word for 'cloud' in Young Avestan was *mayγ-a-* ⟨*maēγa-*⟩, while *aβ-ra-* ⟨*aβra*⟩ had the meaning 'rain cloud'. However, in other Iranian languages descendants of both these stems can mean 'cloud' in general. So, it is reasonable to think that *ab-ra-* and *mayg-a-* were semantically very close to each other in Proto-Iranian, and we are forced to treat them as synonyms. In addition, there is a Gathic word *dwa:n-man-* ⟨*duuąnman-*⟩ 'cloud'; it is derived from the root *dwanH-* 'to fume, fly up' (compare Old Indian ⟨*dhvan^i-*⟩ 'to smoke') and could be a synonym of 'cloud' in poetical language.

Proto-Samoyed. *tiɔ̆* (Janhunen 1977: 162) is retained in all Samoyed languages.

Etymological notes. The widespread stem *neb^h-es-* 'cloud' is a substantival derivative from the verbal root *neb^h-* 'to moisten (*vel sim*.)' reconstructed on the basis of Iranian data. The cognate Iranian deverbative stem *ab-ra-* (< *n̥b^h-ro-*) 'cloud' represents a different morphological pattern, thus we mark Iranian *ab-ra-* with a separate index in the derivational drift-free dataset (Stage-2).

15. 'cold (adj.)'.

Old Irish. *uar* ⟨*úar*⟩.

Proto-Brittonic. *oyr*, cognate with Old Irish ⟨*úar*⟩, is retained in Welsh. The word for 'cold' in Cornish and Breton is derived from the word for 'ice'.

Proto-Germanic. *kald-a-* is retained in all languages.

Proto-Slavic. There are two candidates intertwined within the subgroups: (1) *xald-in-* ⟨*xoldьn-*⟩, and (2) *sto:d-en-* ⟨*studen-*⟩. Out of them, *xald-in-* means either generally 'cold' or specifically 'cool, fresh' in individual lects. Since the deriving substantive *xald-u* ⟨*xoldъ*⟩ is to be reconstructed with the meaning 'coolness, freshness' (as attested in all three subgroups), it is natural to reconstruct the adjectival meaning 'cool, fresh' for *xald-*

30

in-. On the contrary, *stoːd-en-* normally means 'cold' (or 'very cold') in attested lects, having been derived from the substantive *stoːd- ⟨*studъ ~ *studь ~ *studa⟩* which is likely to be reconstructed with the substantive meaning 'cold'. Thus we fill the slot with *stoːd-en-*.

Proto-East Baltic. *šal-ta-* is retained in both Lithuanian and Latvian as well as in Old Prussian.

Proto-Iranian. *car-ta-* is retained in Avestan and the majority of Middle and Modern Iranian languages of all subgroups.

Proto-Samoyed. *täksV- ~ *tätsV- ~ *täçsV- ~ *tässV-* (Janhunen 1977: 159) is retained in all Samoyed languages.

16. 'to come'.

Old Irish. There are two basic verbs with this meaning: *do-ˈig ⟨do-icc⟩* and *do-tʲʲeːd ⟨do-tét⟩* (derived with the prefix ⟨do-⟩ from a suppletive verb ⟨téid⟩ 'to go'). According to DIL, ⟨do-icc⟩ replaces ⟨do-tét⟩ in later versions of early texts and the Modern Irish verb for 'to come' continues ⟨do-icc⟩. The original semantic difference between the two verbs is unclear, but ⟨do-icc⟩ is usually used with direct object denoting the goal of movement. One of the suppletive stems of *do-tʲʲeːd ⟨do-tét⟩* is preterite *do-lˈuðʲ ⟨do-luid⟩*, cognate to Greek ⟨ἐλυθ-⟩.

Proto-Brittonic. *do-aɣ-*, derived with the prefix *do-* from the verb 'to go', is retained in all languages. The verb has a suppletive verbal noun *do-vod*, derived from *bod* 'to be'.

Proto-Germanic. *kwem-an-* is retained in all languages.

Proto-Slavic. The suppletive pair *priː=xad-iː- ⟨*prixoditi⟩* [imperfective] / *priː=yiː-d- ⟨*priti⟩* [perfective] can be safely reconstructed. These stems are derived from *xad-iː- ⟨*xodi-ti⟩* 'to walk' and *iː-d- ⟨*id-, *i-ti⟩* 'to go' respectively with help of the directional prefix *pri-* 'to'.

Proto-East Baltic. Provisionally we fill the slot with *at=ey-*, a prefixed derivative from *ey-* 'to go', which is attested with the meaning 'to come' in Lithuanian, but 'to go aside' in Latvian. The stem *at=ey-* corresponds to the common Indo-European pattern "directional prefix + to go" for 'to come' and thus has a good chance to represent a proto-term, cf. the same pattern although with a different prefix in Old Prussian *per=ey-* 'to come'. The second candidate is *naːk-* meaning 'to rip' in Lithuanian and 'to come' in Latvian.

Proto-Iranian. Avestan prefixless *gam-* ⟨gam-⟩ 'to come' has Old Indic cognates with the same meaning and thus should reflect the original situation. In middle and modern Iranian languages, the meaning 'to come' is usually expressed by prefixed verbs some of which contain the same root **gam-*.**Proto-Samoyed**. **toy-* ~ **tuy-* (Janhunen 1977: 164), retained in all Samoyed languages, goes back to Proto-Uralic **tuli-* 'to come'.

Etymological notes. A criss-crossed configuration with two kinds of form involved. (1) The verb **gʷem-* meaning 'to come' in Tocharian, Armenian, Latin, Germanic, Indo-Iranian. (2) The verb **ey-* 'to go' modified with various spatial prefixes: Anatolian, Baltic, Slavic, as well as in some other Indo-European languages outside our database, e.g., in Modern Iranian lects. The verb **gʷem-* has the advantage due to its morphological primariness, whereas the pattern 'to go' plus directional prefixes most likely represents parallel introductions. Thus we mark the Anatolian, Baltic and Slavic forms with three different indexes in the homoplasy-optimized dataset (Stage-3). Moreover, there is an etymological match between Ancient Greek *eltʰ-* ⟨ἐλθ-⟩ and Old Irish *do=l:'uðʸ* ⟨do-luid⟩, both roots function as a suppletive aorist. This coincidence can be due to chance, but on the other hand, it is not excluded that the Greek-Irish match reflects an old suppletive paradigm. Because of this, just as in the similar case with 'to go' (q.v.), we keep the same index for *eltʰ-* and *do=l:'uðʸ* in the homoplasy-optimized dataset (Stage-3).

17. 'to die'.

Old Irish. *ad-v'alʸ:* ⟨at-baill⟩. The verb has a suppletive preterite *að-b'aθ* ⟨ad-bath⟩. Morphologically, the form ⟨at-baill⟩ consists of a prefix ⟨ad-⟩ or ⟨as-⟩, a "petrified" infixed 3 sg. neuter object pronoun, and a stem ⟨ball⟩ (present tense) / ⟨bel⟩ (other forms). Thus, the original meaning was something like 'he X-s it to ⟨ad-⟩' or 'he X-s it out of ⟨as-⟩', where X is the meaning of ⟨ball⟩, and 'it' refers to life or soul. The most probable etymology connects ⟨ball / bel⟩ with the Indo-European verb **gʷelə-* (**gʷelh₁-*) 'to throw'. Another etymology compares it to the IE verb **gʷelə-* (**gʷelH-*) 'to torment; to sting'. In view of the morphological structure of the Irish verb, the latter etymology seems less likely. The suppletive preterite ⟨ad-bath⟩ goes back to the IE verbal adjective **gʷə-to-* (**gʷh₂-to-*) from the root **gʷaː-* (**gʷeh₂-*) 'to go'.

Proto-Brittonic. **marw-*, a denominative from **marw* 'dead', is retained in all languages.

Proto-Germanic. *dew-an-* / *daw-yan-* may be uncontroversially reconstructed as the principal Germanic verbal root 'die', although its old nominal derivates *daw-θa-* 'death', *daw-da-* 'dead' are consistently retained better than the original verb (in Gothic, for instance, it is the secondary denominal derivate *ga=daw-θ-nan* that is used as the default equivalent for 'to die'). In West Germanic, the verb (but not the nouns) has largely been replaced with euphemisms (e.g. Old English *sweltan*, *steorfan*, etc.) with original meanings such as 'to become weak', 'to suffer / die from hunger (= starve)', etc.

Proto-Slavic. *oː=mer-* ⟨*umerti⟩ is retained in all three subgroups.

Proto-East Baltic. *mir-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *mar-* survives practically in all Iranian languages of all groups and chronological stages; it is one of the most stable Iranian roots. Attested words are often derived from the present stem *mr̥-ya-* or from the past participle *mr̥-ta-*, rarely with prefixes.

Proto-Samoyed. *kvǝ̆-* (Janhunen 1977: 56–57), retained in all Samoyed languages, goes back to Proto-Uralic *kali-* 'to die'.

Etymological notes. Brittonic *marw-* 'to die' represents a denominative verb from the Celtic adjective *mar-wo-* (< *mr̥-wo-*) 'dead', which, in turn, goes back to the common Inner IE verb *mer-* 'to die'. Because of the part of speech change, we mark the Proto-Brittonic form with a separate index in the derivational drift-free dataset (Stage-2).

18. 'dog'.

Old Irish. *ku:* ⟨*cú*⟩.

Proto-Brittonic. *ki:* is retained in all languages.

Proto-Germanic. *xund-a-* is well attested in all ancient and most modern languages.

Proto-Slavic. *pis-u* ⟨*pьsъ⟩ is retained in all three subgroups.

Proto-East Baltic. *šun-* is retained in both Lithuanian and Latvian.

Proto-Iranian. The best candidate is *cwan-* with Indo-Aryan cognates, perfect Indo-European etymology and wide distribution (*cwan-* or its diminutive *cwa-ka-* are attested in Avestan and all Iranian subgroups). The second candidate is *kuta-* / fem. *kuti(:)*, which

is widespread mostly in Eastern Iranian languages and has a weak external etymology. The third stem is *gad-w-a-. Avestan *gað-w-a* ⟨*gaδuua-*⟩ either designates a dog of a special breed or serves as a stylistic synonym for 'dog'; descendants of this root in other Iranian languages often mean 'puppy' (Ossetic, Yaghnobi) or 'a multicolored dog; a dog with a white mark' (Wakhi), 'a dog with cut ears' (Shughni). Attested meanings show that *gad-w-a-* can hardly be a main word for 'dog' in Proto-Iranian.

Proto-Samoyed. *wən* (Janhunen 1977: 173–174) is retained in all daughter languages. This word is possibly a loan from Tocharian (cf. the Tocharian *n*-stem as in Tocharian B acc.sg. *kwen* ⟨*kweṃ*⟩ 'dog').

19. 'to drink'.

Old Irish. *ivʸ-əðʸ* ⟨*ibid*⟩ has a suppletive conjunctive *l:u:s-* ⟨*lús-*⟩ and verbal noun *o:ol* ⟨*óol*⟩.

Proto-Brittonic. *ɪv-* is retained in all languages.

Proto-Germanic. *drenk-an-* is retained in all languages.

Proto-Slavic. *pi:-*, *piy-* ⟨*piti*⟩ is retained in all three subgroups.

Proto-East Baltic. *ger-* is retained in both Lithuanian and Latvian.

Proto-Iranian. In Iranian languages *hwar-* has polysemy 'to eat / to drink' in Avestan, Southwestern languages (Middle Persian, Modern Persian etc.), Northwestern languages (Balochi), Eastern languages (Sogdian, Yaghnobi), Ormuri. Such distribution suggests that *hwar-* could be polysemous already in Proto-Iranian. The second candidate is the root *pi:-* / *pa:-* whose antiquity is proven by the Indo-Aryan cognates. This root forms the verbs for 'to drink' in some languages of the Pamir group (Wakhi ⟨*puv-, pəv-* / *pit-*⟩, Ishkashimi ⟨*pьv-* / *pьvd-*⟩, Sanglechi ⟨*pöv-* / *pöv-*⟩ 'to drink'). Further cf. such cognate forms as modern Persian ⟨*nabīd*⟩ 'wine, date-wine', Kurdish Kurm. ⟨*rāvīn*⟩ 'to dine, taste'. At the current stage of research it is hard to determine whether the Pamir languages possess a retention or represent a backward development. Thus we are forced to reconstruct *hwar-* and *pi:-* / *pa:-* as technical synonyms for Proto-Iranian.

Proto-Samoyed. *ə-r-* (Janhunen 1977: 21–22), retained in Nenets, Mator and Selkup. Replaced in some daughter languages by *witV-*, derived from *wit* 'water'. The Uralic verb

*yɨɣi- ~ *ɨɣi- 'to drink' (> Samoyed *ə-r-) is a likely cognate of Indo-European *eːgʰʷ- (*h₁egʰʷ) 'to drink', see Kassian et al. 2015: 320.

20. 'dry (adj.)'.

Old Irish. *tʲirʲəmʲ* ⟨*tirim*⟩.

Proto-Brittonic. **sɪx*, borrowed from Latin ⟨*siccus*⟩, is retained in all languages.

Proto-Germanic. **θurz-u-* is the best candidate, well attested in all three branches (although in Gothic its semantics is largely limited to 'dried up, withered').

Proto-Slavic. **soːx-* ⟨**sux-*⟩ is retained in all three subgroups.

Proto-East Baltic. **saws-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **huš-ka-* is retained in Avestan and all Iranian subgroups. This root is stable, but sometimes contaminates with a less frequent **hiš-k-u-* 'dry; dried'.

Proto-Samoyed. **kʌsə̆-* (Janhunen 1977: 60–61), retained in Northern Samoyed and Mator, goes back to Proto-Uralic **kosʲki* 'dry'.

Etymological notes. Albanian *θa-tə* is a synchronic participle from a verb 'to dry', which in turn is a denominative from the old adjective **saws-o-* 'dry' (attested as Balto-Slavic **saws-a-* and Indo-Iranian **suš-k-a-*, the latter with a diminutive suffix); because of this we mark the Albanian form with a separate index in the derivational drift-free dataset (Stage-2). Old Irish *tʲirʲəmʲ* and Germanic **θurz-u-* represent different adjectival formations from the verb **ters-* 'to dry (vel sim.)' with the suffix **-i-mi-* in Celtic and the suffix *-u-* in Germanic; because of this we assign two different index to the Old Irish and Germanic forms in the derivational drift-free dataset (Stage-2).

21. 'ear'.

Old Irish. There are two candidates: *kluạs* ⟨*clúas*⟩ 'act of hearing / sense of hearing / ear' (verbal noun of ⟨*ro-cluinethar*⟩ 'to hear') and *aṷ* ⟨*áu*⟩, an old IE word for 'ear'. According to DIL, ⟨*áu*⟩ is "early obsolete and confined to heroic lit. and poetry". However, an examination of textual examples in DIL suggests that the difference between the two words was

semantic rather than stylistic. Contexts where ⟨*clúas*⟩ is used frequently have to do with ear as an instrument of hearing, while ⟨*áu*⟩ is mostly used in "anatomic" contexts, such as 'it extends from my ankle to my ear' (of a shield), 'from ear to tail' (of a bull), 'he bared his teeth as far as his ears', 'with large ears', 'white cows with red ears' etc. Especially interesting is a context where both words co-occur: *di cluais mo dhá ó* 'the hearing of my ears'. Since the GLD standard requires rather "anatomic" than "functional" word for 'ear', we choose *au̯* ⟨*áu*⟩ for the wordlist.

Proto-Brittonic. **skovarn*, retained as the main word for 'ear' in Cornish and Breton. A Welsh reflex also exists, but the main Welsh word for 'ear' is ⟨*clust*⟩ < Proto-Brittonic **klü:st*, cognate to Old Irish ⟨*clúas*⟩. Since the original meaning of **klü:st* was 'hearing' (cf. Old Irish), we choose **skovarn* for the wordlist.

Proto-Germanic. **awz-o:n* is retained in all languages.

Proto-Slavic. **o:x-a* ⟨**uxo*⟩ is retained in all three subgroups.

Proto-East Baltic. **aws-i-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **gawš-a-* is retained in all subgroups and is very stable in the majority of Iranian languages. Derived from the verbal root **gawš-* 'to listen / to perceive'; the related Old Indian verb ⟨*ghoṣ-*⟩ means 'to sound; to cry loudly'. The root **uš-* with a more solid Indo-European etymology survives only in Avestan and Middle Iranian languages mostly with the sense 'an ability to hear and understand', 'reason', 'sense', 'memory' etc. Avestan *kar-na-* ⟨*karəna-*⟩ 'ear (of daevic creatures)' does not have descendants with this meaning in other Iranian languages.

Proto-Samoyed. **kʊw* (Janhunen 1977: 62) is retained in all Samoyed languages.

22. 'earth'.

Old Irish. *talə̃ṽ* ⟨*talam*⟩.

Proto-Brittonic. **dɪyar* is retained in all languages.

Proto-Germanic. **erθ-o:* is retained in all languages.

Proto-Slavic. **zem-i* ⟨**zemь*⟩ and especially its derivative **zem-y-a:* ⟨**zemja*⟩ are retained in all three subgroups.

Proto-East Baltic. **žem-y-a:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **ʒam-* with different suffixes is the most widespread root in all subgroups of Iranian languages. It has a perfect Indo-European etymology. Gathic has the word *bu:-mi-* ⟨*būmi-*⟩ 'earth (opposed to heaven) / land' < Iranian **bu:-mi-* / **bu:-mi:-* with Indian cognates, but there are no instances with the meaning 'earth (soil)' for ⟨*būmi-*⟩. This stem survives in all subgroups predominantly in the Middle Iranian period, often with the meaning 'land', 'world' or 'ground'; it is also the generic designation of 'earth' in Ormuri. According to its semantics in Old and Middle Iranian languages, **bu:-mi-* / **bu:-mi:-* could not be the main word for 'earth' in Proto-Iranian.

Proto-Samoyed. **yvǯ* (Janhunen 1977: 36–37) is retained in the meaning 'earth (soil)' in Northern Samoyed, Mator and Kamass. The root apparently also meant 'sand', q.v.

23. 'to eat'.

Old Irish. *iθ^y-əð^y* ⟨*ithid*⟩, a suppletive verb. The conjunctive ⟨*ess-*⟩ and perfect ⟨*do-fúaid*⟩ stems go back to the Proto-Celtic root **ed-*, while the present stem is either a denominative from Proto-Celtic **i-tu-* 'food' (< PIE **pi-tu-*), or a result of contamination of **ed-* and **itu-*.

Proto-Brittonic. **dıbr-* is retained in Cornish and Breton. Welsh has ⟨*bwyta*⟩, a denominative from ⟨*bwyd*⟩ 'food'. Note, however, that Welsh also has ⟨*ysu*⟩ 'to consume (food), eat, devour'. Although not the main word for 'to eat', this verb possibly reflects Proto-Celtic **ed-*.

Proto-Germanic. **et-an-* is well represented in all three branches, although the root may have already begun to form negative / vulgar connotations ('to eat /of animals/', 'to eat greedily') in Proto-Germanic, since in Gothic and a number of Scandinavian languages it is only preserved in this "vulgar" meaning.

Proto-Slavic. **e:d-* ⟨**ěd-*⟩ is retained in all three subgroups.

Proto-East Baltic. **e:d-* is retained as a generic term for 'to eat' in Latvian, having narrowed its meaning to 'to gobble, guzzle' in Lithuanian.

Proto-Iranian. **hwar-* is retained in Avestan and all Iranian subgroups. This root has no reliable IE cognates. See notes to 'to drink' for the probable Proto-Iranian polysemy 'to drink / to eat'.

Proto-Samoyed. *ə̆m-* (Janhunen 1977: 15), *por-* (Janhunen 1977: 127–128). Proto-Samoyed apparently had two verbs for 'to eat'. The first, *ə̆m-* (< Proto-Uralic *imi-* 'to suck'), is preserved in most daughter languages, while the second, *por-* (< Proto-Uralic *puri-* 'to gnaw, to bite') is retained only in Mator. However, derivatives from both roots — *ə̆m-sə* 'meat, food' and *por-sə* 'fish flour' can be safely reconstructed for Proto-Samoyed. This fact suggests that both *ə̆m-* and *por-* meant 'to eat' in Proto-Samoyed, possibly depending on the type of food.

Etymological notes. Armenian and Albanian form the suppletive preterite with help of the root *gʷerə-* (*gʷerh₃-*) that represents a transparent parallel development 'to devour' > 'to eat'. We mark Armenian perfective *kɛɹ-ay* and Albanian aorist *hən=gra* ⟨*hëngra*⟩ with two separate indexes in the homoplasy-optimized dataset (Stage-3).

24. 'egg'.

Old Irish. *oɣ* ⟨*og*⟩ (< *ug-es-* or *og-es-*) is sometimes viewed as an irregular reflex of the PIE word for 'egg'. However, the irregularity is too great: the only thing that these words have in common is the vocalic anlaut. Neither the root vowel, nor consonant *ɣ* can reflect IE protoform (whether it is reconstructed as *oːwyom* or *oːyom*). Thus, we code this word as not cognate to words for 'egg' in other IE languages.

Proto-Brittonic. *ɔːy*, retained in all languages, is a regular continuation of the main PIE word for 'egg'.

Proto-Germanic. *ayy-a-* is retained in all languages (in modern English, re-borrowed from Scandinavian).

Proto-Slavic. *aːy-e, *aːy-ik-a* ⟨*aje, *ajьko*⟩ is retained in all three subgroups.

Proto-East Baltic. *paut-a-* is retained in both Lithuanian and Latvian as well as in Old Prussian. In modern Lithuanian and Latvian, it tends to be superseded with new formations from the roots 'skull, shell' and 'round' respectively.

Proto-Iranian. *aːy-a- ~ *aːy-a-ka-* is retained in Avestan and several Iranian languages belonging to different subgroups (Khotanese, Kurdish, Parachi, Ossetic etc.); external cognates point to the original meaning 'egg'. Another candidate, *tawx-man-*, clearly had the original meaning 'seed / embryo / kinship'.

Proto-Samoyed. *mǎnʊ* (Janhunen 1977: 86), retained in Enets, Nganasan and Kamass (Selkup reflex of this word means 'penis'), goes back to Proto-Uralic *muna* 'egg'.

25. 'eye'.

Old Hittite. The etymological connection between *saguw-a-* and Nuclear IE *okʷ-* (*h₃ekʷ-*) 'eye' is widely accepted among modern Indo-Europeanists, although there is no consensus about phonological details of the reconstruction.

Old Irish. *suːlʸ* ⟨*súil*⟩.

Proto-Brittonic. *lːugad* is retained in all languages.

Proto-Germanic. *awg-oːn* is retained in all languages.

Proto-Slavic. *ak-a* ⟨*oko*⟩ is retained in all three subgroups.

Proto-East Baltic. *ak-i-* is retained in both Lithuanian and Latvian.

Proto-Iranian. The deverbative *čaš-man-* is retained in Avestan and the majority of Iranian languages belonging to all subgroups; one of the most stable items. Avestan dual. *aš-i-* ⟨*aš-i-*⟩ 'eyes (of Daevic creatures)' has a perfect Indo-European etymology, but represents only a relic already in the earliest texts.

Proto-Samoyed. *sǎymä* (Janhunen 1977: 132), retained in all daughter languages, goes back to Proto-Uralic *silmä* 'eye'.

26. 'fat'.

Old Irish. *iːθ* ⟨*íth*⟩.

Proto-Brittonic. *bloneg* is retained in all daughter languages, although apparently it is not the main word with this meaning in Welsh and Breton. Welsh ⟨*braster*⟩ 'fat' (n.) is derived from ⟨*bras*⟩ 'fat, thick' (adj.); Breton has ⟨*lard*⟩, borrowed from French; another French loanword is Welsh ⟨*saim*⟩, Cornish ⟨*seym*⟩. Middle Irish ⟨*blonac*⟩ 'fat, lard, grease' is apparently borrowed from Brittonic.

Proto-Germanic. In all languages where it is attested, the noun 'fat' is formed from the adjectival / verbal root *f(a)yt- '(to be) fat, fattened': Proto-Scandinavian *fit-oːn- ~ *fit-iːn-, Proto-West Germanic *fayt-ida- (Kroonen 2013: 124, 142). No proper equivalent, however, is attested in Gothic, and the presence of several morphological variants somewhat undermines the Proto-Germanic reconstruction. Technically, however, the root *f(a)yt- may be entered into comparison, since it forms the basis for the Swadesh item in two out of three primary branches.

Proto-Slavic. Despite general instability of this item, the Proto-Slavic term can be safely reconstructed as *toːk-u ⟨*tukъ⟩. This stem is retained in West languages as well as in Old Church Slavonic and Old Russian, moreover it has Baltic cognates with the same meanings.

Proto-East Baltic. *tauk-a- is retained in both Lithuanian and Latvian.

Proto-Iranian. There are several candidates, but *piː-wah- is retained in Avestan and many Middle and Modern Iranian languages from all Iranian subgroups, and has reliable Indo-Aryan and Indo-European cognates.

Proto-Samoyed. *yür (Janhunen 1977: 50), *tuyt ~ *çuyt ~ *tuyç ~ *çuyç (Janhunen 1977: 166). The semantic difference between the two Proto-Samoyed words for 'fat' is not clear. The first one, *yür, is a Turkic loanword, the second one probably is the original Samoyed word for 'fat'.

Etymological notes. Ancient Greek piː-mel-éː, Old Irish iːθ (< *piː-tu-) and Indo-Iranian *piː-was represent different nominal derivatives from the verbal root *peyə- (*peyH-) 'to swell (*vel sim*.)'. Because of this we mark the Greek, Irish and Indo-Iranian forms with separate indexes in the derivational drift-free dataset (Stage-2). It is not excluded that Latin $ping^w$-e and Germanic *fayt- ~ *fit- also contain the same root, but we prefer to treat the Latin and Germanic forms as independent.

27. 'feather'.

Old Irish. $et^y e$ ⟨ette ~ eitte⟩. Polysemy: 'wing / fin / feather, plume'. An alternative candidate is kluːṽ ⟨clúm⟩ 'feathers, plumage, down' (borrowed from Latin ⟨plūma⟩). In modern Goidelic languages the descendants of these words differ in the following way: Modern Irish ⟨clúmh⟩ '(of birds) down, feathers; hair (on body); down (on cheeks); (of animals) fur, coat'

vs. ⟨*cleite*⟩ (anlaut ⟨*cl-*⟩ due to contamination with ⟨*clúmh*⟩) 'feather; quill; plume'; Modern Scottish Gaelic ⟨*clùimh*⟩ 'wool; down (feathers); moult' vs. ⟨*ite*⟩ 'feather, plume; fin'. According to the GLD semantic standard, we choose the word for a single feather (admittedly not very frequent in Old Irish texts in this meaning) rather than a collective word for 'down, feathers'. *et^ʲe* ⟨*ette ~ eitte*⟩ is usually compared to PIE root **pet-* 'to fly', although phonetic details are unclear.

Proto-Brittonic. **plü:ṽ*, borrowed from Latin ⟨*plūma*⟩ is retained in all languages.

Proto-Germanic. **feθr-o:* is retained in all languages.

Proto-Slavic. **per-a* ⟨**pero*⟩ is retained in all three subgroups.

Proto-East Baltic. In both languages new formations are used: **sn-*deverbative from **pluk-* 'to pluck, plume' in Lithuanian and **spalv-a:* in Latvian whose original meaning was something like 'fringe-like'. The choice is irrelevant to phylogenetic purposes. The Latvian stem seems slightly more preferable due to its non-derivative nature.

Proto-Iranian. **par-na-* is retained in Avestan and the majority of Middle and Modern Iranian languages belonging to all Iranian subgroups. The word has direct Indo-Aryan cognates.

Proto-Samoyed. **tuð* (Janhunen 1977: 166), retained in Northern Samoyed, Mator and Selkup, is related to Finno-Ugric **tulka* 'feather / wing'.

Etymological notes. Hittite *patt-ar*, Ancient Greek *pter-ó-n*, Latin *pˈen:-a*, Germanic **feθr-o:* go back to an old heteroclitic noun derived from the verbal root **pet-* 'to fly'. In its turn, Old Irish *et^ʲe* represents another suffixed deverbative from the same root (morphological details are not entirely clear, however), so we mark the Old Irish form with a different index in the derivational drift-free dataset (Stage-2). Tocharian B *par-w-a*, Slavic **per-a* and Indo-Iranian **par-na-* are separate nominal derivatives from the verbal root **per-* 'to move (vel sim.)', so we mark these forms with separate indexes in the derivational drift-free dataset (Stage-2).

28. 'fire'.

Archaic Latin. Latin *ign-is* 'fire' is opposed to the different term for 'fire' in another Italic language: Umbrian *pir*, obl. *pur-* 'fire'.

Old Irish. *tʸenʸe* ⟨*teine*⟩.

Proto-Brittonic. **tan* is retained in all languages.

Proto-Germanic. The root **fuː-* is present in all three primary branches (although already in Proto-Scandinavian / Old Norse it was replaced by the innovation **ailida-*), but with different morphological variants: **fuː-r* in Scandinavian and West Germanic, **fu-n-oːn* in Gothic, with the **-n*-stem variant also preserved in Old Norse *funi* 'flame' (cf. similar situations, but with different distribution, for 'sun' and 'water').

Proto-Slavic. **agn-i* ⟨**ognь*⟩ is retained in all three subgroups.

Proto-East Baltic. **ungn-i-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **aː-tar-* is retained in all subgroups. It is derived from the PIE verbal root **aː-* (**h₂eh₁-*) 'to be hot'.

Proto-Samoyed. **tuɣ* (Janhunen 1977: 166), retained in all Samoyed languages, goes back to Proto-Uralic **tuli* 'fire'.

Etymological notes. IE **ŋg-n-i-* denotes 'fire' in Latin and the Balto-Slavic–Indo-Iranian clade (with the further replacement with **aː-tar-* in Proto-Iranian). This distribution violates the topology of the IE tree obtained at Stage 2 (Fig. 1 in the main text). Note that actually Latin *igni-s* is not even likely to represent a Proto-Italic term for 'fire'. We mark the Latin and Balto-Slavic–Indo-Iranian forms with two separate indexes in the homoplasy-optimized dataset (Stage-3).

29. 'fish'.

Old Irish. *i̯ask* ⟨*íasc*⟩.

Proto-Brittonic. **pɪsk*, borrowed from Latin ⟨*piscis*⟩, is retained in all languages.

Proto-Germanic. **fisk-a-* is retained in all languages.

Proto-Slavic. **ruːb-aː* ⟨**ryba*⟩ is retained in all three subgroups.

Proto-East Baltic. **žuv-i-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *macy-a-* (< *matsy-a*) is retained in Avestan, Southwestern (Middle and Modern Persian), Northwestern (Parthian, Kurdish etc.) and some Eastern Iranian languages (Parachi, Pashto); moreover, it has clear Indo-Aryan cognates. Common Eastern Iranian *kapa-* 'fish' does not occur in other subgroups and has no reliable etymology.

Proto-Samoyed. *kвlä* (Janhunen 1977: 59), retained in all Samoyed languages, goes back to Proto-Uralic *kala* 'fish'.

30. 'to fly'.

Old Irish. *fo=l:ˈu-ə-θər* ⟨fo-luathar⟩.

Proto-Brittonic. *(h)ed-* is retained in Welsh. Goes back to PIE *pet-* 'to fly'; the origin of *h-* in Welsh is unclear. Cornish ⟨nyidzha⟩ and Breton ⟨nijal⟩ are related to Welsh ⟨neidio⟩ 'to jump, leap' — a denominative verb, derived from ⟨naid⟩ 'a leap, jump'.

Proto-Germanic. *flewg-an-* is retained throughout Scandinavian and West Germanic, but the word is not attested in Gothic.

Proto-Slavic. *let-e:-* ⟨*letěti*, *letati*⟩ is retained in all three subgroups.

Proto-East Baltic. An unclear case with three competing stems. First, *skrid-*, attested only in Lithuanian, is a basic verb for 'to fly' in modern Lithuanian. It is possibly a derivative from Baltic *skrey-* 'to be flying; to move round' (also attested in Lithuanian only). Second, *li:d-o:-*, a basic verb for 'to fly' in modern Latvian, is not attested elsewhere. It is apparently an iterative from *lid-* 'to crawl' (also attested in Latvian only). Third, *lek-*, which means 'to fly; to run; to fall' in Lithuanian, but 'to jump' and rarely 'to fly' in Latvian. The choice is irrelevant to phylogenetic purposes. The stem *lek-* seems more preferable since it is attested with the semantics 'to fly' in both East Baltic languages, although neither in Lithuanian nor in Latvian it is a basic expression for this meaning.

Proto-Iranian. *pat-* still preserves its original meaning in several Avestan passages, and the direct Indo-Aryan cognates speak in favor of its antiquity. In the majority of Iranian languages it was superseded by verbs with different semantics (often prefixed) or denominatives from *par-na-* 'feather'.

Proto-Samoyed. *tey-* (Janhunen 1977: 161–162), retained in all languages save Kamass, goes back to Proto-Uralic *şelki-* 'to fly'.

43

31. 'foot'.

Old Irish. *kos* ⟨*cos*⟩. Polysemy: 'foot, leg (of human beings and animals); stem, support, handle, shaft of various objects'.

Proto-Brittonic. **troɣed* is retained in all languages.

Proto-Germanic. **foːt-u-* 'foot' is retained in all languages.

Proto-Slavic. **nag-aː* ⟨**noga*⟩ 'foot / leg' is retained in all three subgroups.

Proto-East Baltic. **kaːɣ-aː* 'foot / leg' is retained in both Lithuanian and Latvian.

Proto-Iranian. **pad-* is one of the most stable Iranian stems with direct Indo-Aryan and IE cognates.

Proto-Samoyed. **vy* (Janhunen 1977: 17) is retained in all daughter languages save Selkup.

32. 'full'.

Old Irish. *lːaːn* ⟨*lán*⟩.

Proto-Brittonic. **lːɔːn* is retained in all languages.

Proto-Germanic. **full-a-* is retained in all languages.

Proto-Slavic. **pil-n-* ⟨**pьln-*⟩ is retained in all three subgroups.

Proto-East Baltic. **pil-n-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **pər-na-* is retained in Avestan and all Iranian subgroups and has direct Indo-Aryan cognates. We follow Cantera (2001) in reconstructing a specific Proto-Iranian combination **ər* as an outcome of PIE long **r̥ː* (**r̥H*) in the position after a labial consonant.**Proto-Samoyed**. **tärðˇ* (Janhunen 1977: 158), retained in Selkup as the word for 'full' (Northern Samoyed cognates mean 'interior'), goes back to Proto-Uralic **täwði* 'full' (Aikio 2002: 31–34).

Etymological notes. The variety of forms represents separate adjectival derivatives from the verb **pleː-* (**pleh₁-*) 'to be full'. (1) The **ro*-suffix in Ancient Greek *pléː-reː-*. (2) The **yo-*

suffix in Armenian *li* (the Armenian stem may go back to a stem with the **to*-suffix as well, but the direct Ancient Greek cognate ⟨πλέος⟩ 'full' < **pleː-yo-* makes this solution less natural). (3) The **to*-suffix in Albanian *plɔ-tə*. (4) The **no*-suffix in Latin *plˈeːn-us*, Celtic **lːaːno-*, German **full-a-*, Balto-Slavic **pil-n-a-*, Indo-Iranian **pr̩ː-na-*. We assign four different indexes to these groups in the derivational drift-free dataset (Stage-2).

33. 'to give'.

Old Hittite. We follow modern analysis (Kloekhorst 2008: 615; Puhvel 2011: 56) and reject the old idea that Anatolian **pay-* / **pi-* 'to give' represents contraction of the Anatolian prefix **pe-* 'from' and the verbal root **ay-* whose bare form is attested as Tocharian **ay-* 'to give'. Internal Anatolian evidence suggests that **pay-* / **pi-* is a primary verbal root, although its further IE etymology is vague.

Old Irish. *do=bˈerʸ* ⟨do-beir⟩. Polysemy: 'give, place; bring, get'. The verb has suppletive perfect stems: *do=rˈːad* ⟨do-rat⟩ in the meaning 'give, place' and *do=ˈug* ⟨do-ucc⟩ in the meaning 'bring, get'. The root of *do=bˈerʸ* ⟨do-beir⟩ reflects PIE **bʰer-* 'to carry'. The suppletive perfect *do=rˈːad* ⟨do-rat⟩ goes back to **to-ro-ad-d-*, where the root is **-d-*, going back either to PIE **doː-* (**deh₃-*) 'to give' or to **dʰeː-* (**dʰeh₁-*) 'to place'.

Proto-Brittonic. **ro=ð-* is retained in all languages. The root **=ð-* can go back either to PIE **doː-* (**deh₃-*) 'to give' or to **dʰeː-* (**dʰeh₁-*) 'to place'. According to S. Schumacher (2000: 131), the verbal noun **roði:ṽ* < **rodi:ma:* speaks in favor of **dʰeː-* (**dʰeh₁-*): **dʰeː-ma:* (**-dʰeh₁-meh₂*) > **-di:-ma:*. Later, **-di:-ma:* was reinterpreted as **-d-i:ma:* and **-i:ma:* spread to other verbs that originally did not have the root **-di:*. We regard this scenario as too speculative, and prefer to mark **ro=ð-* as a reflex of PIE **doː-* (**deh₃-*) in accordance with its semantics.

Proto-Germanic. **geb-an-* is retained in all ancient and modern languages (in modern English, only in the "Scandinavized" variant ⟨give⟩).

Proto-Slavic. **da:-* ⟨**dati*⟩ is retained in all three subgroups.

Proto-East Baltic. **do:-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **da:-* is retained in Avestan and the majority of Iranian languages belonging to all subgroups; this root has direct Indo-Aryan cognates and PIE etymology.

Proto-Samoyed. **mi-* (Janhunen 1977: 94), **tʋ-* (Janhunen 1977: 94). Judging by the situation in Tundra Nenets, the two verbs for 'give' were used depending on the person of recipient: **mi-* (< Proto-Uralic **meɣi-* 'to give') was used with 2nd and 3rd person recipient, while **tʋ-* was required in sentences with 1st person recipient. The Proto-Uralic verb **toɣi-* 'to give' (> Samoyed **tʋ-*) is a likely cognate of Indo-European **do:-* (**deh₃-*) 'to give' (Illich-Svitych 1967: 338), note **-ɣ-* as a counterpart of the IE "laryngeal".

34. 'good'.

Old Irish. *maθʸ* ⟨*maith*⟩.

Proto-Brittonic. **daɣ* and **mad* are both retained in all languages, although the former is the main word for 'good' in Welsh and Cornish and the latter in Breton. Goidelic comparanda also do not solve the question: Old Irish has *maθʸ* ⟨*maith*⟩ as an independent adjective, but *daɣ-* ⟨*dag-*⟩ as a first part of compounds. This distribution cannot be projected to Proto-Celtic, since in Brittonic both **daɣ* and **mad* function as independent adjectives.

Proto-Germanic. **go:ð-a-* is retained in all languages.

Proto-Slavic. **dab-r-* ⟨**dobr-*⟩ is retained in all three subgroups.

Proto-East Baltic. **lab-a-* means generic 'good' in Latvian and 'kind; appropriate' in Lithuanian; the antiquity of **lab-a-* is proven by its meaning 'good' in Old Prussian.

Proto-Iranian. **wah-u-* is retained in Avestan and some other Iranian languages (among them, Khotanese); it has direct Indo-Aryan and Indo-European cognates.

Proto-Samoyed. **sǝ̆mʋ* (Janhunen 1977: 132–133) is retained in Northern Samoyed and Selkup.

35. 'green'.

Old Irish. *glas* ⟨*glas*⟩ is a color term encompassing shades of green, blue and grey. The more specific term *uan̥ʸe* ⟨*úaine*⟩ 'green' is usually applied to the 'artificial' color of clothes, shields etc.

Proto-Brittonic. **glas* is retained in all languages. Brittonic has the same kind of opposition as in Irish: **glas* denotes 'natural' green, blue and grey, while **gwɪrð* (borrowed from Latin) refers to 'artificial' green.

Proto-Germanic. **groːn-i-* is retained in all languages.

Proto-Slavic. **zel-en-* ⟨**zelen-*⟩ is retained in all three subgroups.

Proto-East Baltic. **žal-ya-* is retained in both Lithuanian and Latvian.

Proto-Iranian. It is reasonable to reconstruct **ʒar-i-* with polysemy 'green / yellow'; these meanings are found in Avestan and coincide with the Old Indian situation. In the majority of Iranian languages, this adjective with the meaning 'green' was superseded by **capa-či-* 'green' from **capa-* 'grass, greenery'.

Proto-Samoyed. **tǎŋkV-* ~ **çǒŋkV-* 'green / blue' (Helimski 1997: 356) is retained in Forest Nenets and Mator.

36. 'hair'.

Old Irish. *folt* ⟨*folt*⟩.

Proto-Brittonic. **gwolt* is retained in Welsh, where ⟨*gwallt*⟩ means 'head hair' as opposed to ⟨*blew*⟩ 'body hair'. Cornish and Breton have **blew* in both meanings.

Proto-Germanic. **xeːr-a-* is probably the optimal candidate, since it is retained in all ancient and modern Scandinavian and West Germanic languages. However, in Gothic the equivalent is *tagl* < PG **tagl-a-*, meaning 'hair of the tail' in Scandinavian and West Germanic sources (and then shifting to 'tail' proper in English). There are no decisive arguments on whether it is preferable to assume a narrowing or a broadening of the meaning, so, ultimately, both words may be entered into comparison.

Proto-Slavic. **vals-u* ⟨**volsъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **mat-a-* is attested only in Latvian, it remains etymologically obscure, but nevertheless this is the best candidate for the Proto-East Baltic status. The second and even weaker candidate is **plauk-a-* ~ **plauk-aː*, meaning 'hair' in Lithuanian and 'flock(s)' in Latvian, the Latvian meaning is apparently primary since the Germanic cognates of the Baltic stems have the same semantics 'flock, down'.

Old Indic. We follow Lubotsky 2001: 302, 311 and treat Old Indic ⟨*keśa-*⟩ 'hair' as well as Avestan ⟨*gaēsa-*⟩ 'curls' with its further Iranian cognates as loans from an unknown source.

Proto-Iranian. **warc-a-* is retained in this meaning in Avestan, in Northwestern Iranian languages (Pahlavi) and in several Eastern Iranian languages of different periods (for example, Khotanese, Sogdian, and Pashto). Moreover, it has an Old Indian cognate *w'alſ-a-* ⟨*válśa-*⟩ 'sprout / branch, twig' and Proto-Slavic **vals-u* ⟨**volsъ*⟩ 'hair'. In Western Iranian it has altered meanings, e.g., Persian ⟨*gors*⟩ 'curl'. The stem **maud-a-* 'hair' does not have an etymology and is found only in Southwestern and Northwestern Iranian languages of middle and modern periods. The stem **gawn-a-* originally meant 'body hair (of animals) / color (of horse)'.

Proto-Samoyed. **əptӗ* (Janhunen 1977: 21), retained in all Samoyed languages, goes back to Proto-Uralic **ïpti* 'hair'.

37. 'hand'.

Old Irish. *l:a:ṽ* ⟨*lám*⟩.

Proto-Brittonic. **l:ɔ:ṽ* is retained in Welsh, Cornish and Old Breton. Replaced in Modern Breton by ⟨*dorn*⟩ < **du/orn* 'fist'.

Proto-Germanic. **xand-u-* is retained in all languages.

Proto-Slavic. **rank-a:* ⟨**rǫka*⟩ 'hand / arm' is retained in all three subgroups.

Proto-East Baltic. **rank-a:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **ʒas-ta-* (or its dissimilative variant **das-ta-*) is retained in Avestan and the majority of Iranian languages belonging to all subgroups. This stem has direct Indo-Aryan cognates.

Proto-Samoyed. **utʋ* (Janhunen 1977: 30) is retained in all Samoyed languages. Judging by its phonology (the vowel combination), the word can hardly be inherited from Proto-Uralic, so strictly speaking **utʋ* must represent a borrowing from an unknown source.

38. 'head'.

Albanian. *kɔk-ə* ⟨*kokë*⟩ was borrowed from Latin ⟨*coccum*⟩ 'berry'. The attested polysemy 'head / bulb / berry / grain' of the Albanian term indicates that the semantic development 'berry' > 'head' is an internal Albanian evolution.

Old Irish. *kʲen:* ⟨*cenn*⟩.

Proto-Brittonic. **pen:* is retained in all languages.

Proto-Germanic. **xawbid-a-* is retained in all languages, except for German, where it has been replaced by ⟨*Kopf*⟩ (ultimately, of Latin origin).

Proto-Slavic. **galv-a:* ⟨**golva*⟩ is retained in all three subgroups.

Proto-East Baltic. **galv-a:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **car-ah-* is retained in Avestan and the majority of Iranian languages belonging to all subgroups. This stem has direct Indo-Aryan cognates.

Proto-Samoyed. **vɨwʊ* (Janhunen 1977: 17), retained in Northern Samoyed and Mator, goes back to Proto-Uralic **oywa* 'head'.

Etymological notes. There exists a traditional hypothesis about a connection of Armenian *glux* 'head' and Balto-Slavic **galv-a:* 'head', but this comparison faces too many difficulties. The Tocharian A form ⟨*śpāl*⟩ 'head' is worth noting since it can be safely reconstructed as a Proto-Tocharian term for 'head' etymologically matching Ancient Greek *kepʰal-é:* ⟨κεφαλή⟩ 'head'.

39. 'to hear'.

Old Irish. *r:o=klʲuˈ-nʲ-ə-θər* ⟨*ro-cluinethar*⟩.

Proto-Brittonic. **kluw-* is retained in all languages.

Proto-Germanic. **xauz-yan-* is retained in all languages.

Proto-Slavic. There are two candidates intertwined within the subgroups: (1) **slu:š-a:-* ⟨**slyšati*⟩, and (2) **čo:-* ⟨**čuti*⟩. Although the exact derivational history of **slu:š-a:-* (< **slu:x-e:-*) remains unclear, this verb seems preferable for the status of the Proto-Slavic term for 'to hear', firstly, because **slu:š-a:-* is used as a basic term in ancient languages: Old Church Slavonic, Old Russian, Old Novgorod, Old Czech, Old Polish, whereas **čo:-* with the mean-

ing 'to hear' is mostly restricted to modern lects. Secondly, *čo:- denotes a wide range of acts of perception (feeling, hearing, tasting and so on), so it is most likely that *čo:- is to be reconstructed simply with the generic meaning 'to feel'. Distinct from *slo:š-a:-ye- ⟨*slušati⟩ 'to listen', derived from the noun *slo:x-u ⟨*sluxъ⟩ '(ability of) hearing'.

Proto-East Baltic. *gird-e:- is retained in both Lithuanian and Latvian.

Proto-Iranian. *craw- is retained in its basic meaning 'to hear' in Avestan, Balochi and Eastern Iranian languages (the Shughni group and Yazghulami); other Iranian languages of all subgroups preserve only derived meanings, like 'to sing', 'to be heard'. This root has direct Indo-Aryan cognates. Verbs with the root *gawš- (mostly prefixed) are also attested as 'to hear' in many Iranian languages belonging to all subgroups. However, Avestan gawš- ⟨gaoš-⟩ means 'to listen / to perceive' and Old Indian ⟨ghoṣ-⟩ means 'to sound'. So, it is unnecessary to treat *craw- and *gawš- as Proto-Iranian synonyms.

Proto-Samoyed. *yüntə̆- (Janhunen 1977: 49) is retained in most Samoyed languages.

Etymological notes. The reliable candidate for the status of the Proto-IE (or at least Proto-Nuclear IE) term is $k^y lew$-(s)- which shifted to more specific meanings 'to hear, understand, obey' and 'to listen' in Ancient Greek and Germanic respectively. Thus the root cognacy between Ancient Greek akú:-o: ⟨ἀκούω⟩ and Germanic *xauz-yan- represents parallel developments in these groups, although the origin and the semantics of the proto-stem in question remain obscure. We mark the Greek and Germanic forms with different indexes in the homoplasy-optimized dataset (Stage-3).

40. 'heart'.

Old Irish. $kr^y i\eth^y e$ ⟨cride⟩.

Proto-Brittonic. *kalon, retained in all languages, is apparently borrowed from a Romance form close to Old French chaudun 'entrails'.

Proto-Germanic. *xert-o:n is retained in all languages.

Proto-Slavic. *sird-ik-o ⟨*sьrdьce⟩ is retained in all three subgroups.

Proto-East Baltic. *šird-i- ~ *šird-a- is retained in both Lithuanian and Latvian.

Proto-Iranian. *ʒr̥d-* and several suffixed derivatives with the same meaning are retained in Avestan and the majority of Iranian languages. This root has direct Indo-Aryan and IE cognates. The Proto-Indo-Iranian form has initial *ʒʰ-* as from IE *gʸʰ-* instead of expected **c-* < IE *kʸ-*, but this irregularity can hardly reject the straightforward IE etymology for the Proto-Indo-Iranian term.

Proto-Samoyed. *säyə̌*, retained in all Samoyed languages, goes back to Proto-Uralic *sʸäðä* 'heart'.

41. 'horn'.

Old Irish. *aðərk* ⟨*adarc*⟩ is possibly borrowed from a language related to Basque, cf. Basque ⟨*adar*⟩ 'horn'.

Proto-Brittonic. *korn*, retained in all languages, is apparently borrowed from Latin ⟨*cornu*⟩.

Proto-Germanic. *xurn-a-* is retained in all languages.

Proto-Slavic. *rag-u* ⟨*rogъ*⟩ is retained in all three subgroups.

Proto-East Baltic. *rag-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *cr-u:-* and derivatives of this root are retained in the majority of Iranian languages of all subgroups. This stem has Indo-Aryan cognates.

Proto-Samoyed. *amtə̌* (Janhunen 1977: 20) is retained in all Samoyed languages.

Etymological notes. The majority of the involved terms for 'horn' represents various suffixal formations from the root *kʸer-* 'head' (Kassian et al. 2015: 331). The initial consonant of Hittite *karaw-ar* definitely goes back to Proto-Anatolian *kʸ-* (< IE palatalized velar), as proven by the cognate Luwian form in *c-* ⟨*z-*⟩. Due to multiple origin possible for the initial consonants in Tocharian *kror-* 'horn', the Tocharian stem can be formally compared with either the bulk of IE forms for 'horn' (if < *kʸ-*) or with Armenian *εłǯεwɹ* 'horn' (if < *gʰ-*). The first solution is much more natural, the Armenian form is thus treated as unrelated. Proto-Tocharian *kror* could represent a direct morphological cognate of Hittite *karaw-ar*.

42. 'I'.

Old Irish. *me: ⟨mé⟩.*

Proto-Brittonic. **mɪ* is retained in all languages.

Proto-Germanic. The suppletive paradigm **ek* [nom.] / **me-* [obl.] is retained in all languages.

Proto-Slavic. The suppletive paradigm **a:z-u* ⟨**azъ*⟩ [nom.] / **me-n-* ⟨**mene*⟩ [obl.] is retained in all three subgroups.

Proto-East Baltic. The suppletive paradigm **eš ~ *aš* [nom.] / **man-* [obl.] is retained in both Lithuanian and Latvian.

Proto-Iranian. The suppletive paradigm **aʒ-am / ma-* is retained in Avestan and several Iranian languages (for example, Parachi and Ossetic); in many languages the form **aʒ-am* has been superseded by the oblique stem *ma-*. The paradigm has direct Indo-Aryan cognates and PIE etymology.

Proto-Samoyed. **mǐ-n* (Janhunen 1977: 86) is retained in all Samoyed languages. The Proto-Uralic personal pronoun **mi-n* 'I' (> Samoyed **mǐ-n*) is a likely cognate of Indo-European **me-* 'I'.

43. 'to kill'.

Old Irish. *marv-əðʸ* ⟨*marbaid*⟩ is unquestionably the main word for 'to kill'. *org-əðʸ* ⟨*orcaid*⟩, listed by Lucht (2007: 226) as the main Old Irish synonym for 'to kill', actually means 'slay'. The search in eDIL yields almost hundred of textual examples where ⟨*marbaid*⟩ is translated as 'kill', but only two or three each for ⟨*orcaid*⟩ and ⟨*gonaid*⟩ (the main meaning of the latter verb is 'to wound'). Search for words translated as 'slay' yields roughly equal number of examples with ⟨*marbaid*⟩, ⟨*orcaid*⟩ and ⟨*gonaid*⟩.

Proto-Brittonic. **l:að-*, related to Old Irish ⟨*slaidid*⟩ 'strike; slay; plunder, destroy' is retained in all languages.

Proto-Germanic. This word is highly unstable in Germanic languages and is seemingly subject to a common process of euphemization. Cf. Old Norse ⟨*drep-a*⟩ 'to strike, to kill' and its cognates in most modern Scandinavian languages < **drep-an-* 'to hit, to strike'; Gothic

⟨us=kwim-an⟩ 'to kill' (literally 'to come out'); Old English ⟨cwell-an⟩ 'to kill' < *kwel-an- 'to suffer, be tortured', etc. Probably the most likely candidate is the verbal stem *daw-d-yan-, a causative formation from *daw-da- 'dead' (see 'die'): it functions as the main equivalent for 'to kill' in several modern languages (Swedish, German) and is sufficiently attested in the same meaning in ancient languages (Gothic ⟨ga-dauþjan⟩ - encountered only once; Old Norse ⟨deyða⟩, etc.) to be definitively reconstructible for Proto-Germanic, although whether it truly functioned in that language as the most frequent and neutral equivalent for 'to kill' is impossible to tell.

Proto-Slavic. The perfective prefixed verb *oː=bi:- ⟨*ubiti⟩ 'to kill', derived from *bi:- ⟨*biti⟩ 'to beat', is retained in South and East subgroups. In West lects, superseded with either *zaː=bi:- ⟨*zabiti⟩ of the same perfective model or with *mar-i:- ⟨*moriti⟩, a causative from *mer- 'to die' q.v.

Proto-East Baltic. *gal-in- is retained in Latvian (as the secondary stem *noː=gal-in-aː-), its antiquity is proven by the same meaning 'to kill' in Old Prussian. Derived from Proto-Baltic *gal-a- 'end'.

Proto-Iranian. *gan- with the polysemy 'to hit / to strike / to slay / to kill' is retained in Avestan and all Iranian subgroups. The verb has direct Indo-Aryan and IE cognates. In many Iranian languages, it preserves only the meanings 'to hit / to strike', but it is the main word for 'to kill' in Avestan, Parachi, Ormuri and many Eastern Iranian languages (Khotanese, Shughni, Ishkashimi etc.). In Northwestern and Southwestern Iranian languages the main verb with this meaning is *kawš-. Considering the facts that this root is found only in one Avestan passage in a prefixed stem fra=kawš- ⟨fra.kaoš-⟩ 'to slay, to kill', and its Indo-Aryan cognates have the meaning 'to tear, to tear to pieces / to tear out', it is natural to think that *kawš- was not the main root for 'to kill' in Proto-Iranian.

Proto-Samoyed. *kʌ̆-tʌ- (Janhunen 1977: 57), retained in all daughter languages, is a causative of Proto-Samoyed *kʌ̆- 'to die' (< Proto-Uralic *kali- 'to die').

44. 'knee'.

Old Irish. gluːn ⟨glún⟩.

Proto-Brittonic. *gliːn is retained in all languages.

Proto-Germanic. **knew-a-* is retained in all ancient and modern languages.

Proto-Slavic. **kal-eːn-a* ⟨**kolěno*⟩ is retained in all three subgroups.

Proto-East Baltic. **kel-ya-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **ʒaːn-u-* is retained in Avestan and the majority of Iranian languages belonging to all subgroups. It has direct Indo-Aryan and PIE cognates.

Proto-Samoyed. **puǯ* (Janhunen 1977: 130), retained in all languages save Kamass, is related to Finno-Ugric words for 'knee'.

Etymological notes. Slavic **kal-eːn-a* ⟨**kolěno*⟩ and Baltic **kel-ya-* look like parallel formations from a Balto-Slavic ablaut root **kel- / *kal-* whose meaning is unknown, but it could hardly be 'knee' *per se*. So we mark the Slavic and Baltic forms with separate indexes in the derivational drift-free dataset (Stage-2). The similarity between Proto-Albanian **gluna* 'knee' and Proto-Celtic **gluːn-os-* 'knee' is very suspicious. Formally it cannot be excluded that we are dealing with parallel formations from the common IE root modified with an unclear *n*-suffix and affected by occasional dissimilation *n-n > l-n*. But the opposite solution is also possible: the Albanian and Celtic forms are exclusive cognates being unrelated to IE **gʸonu* 'knee'.

45. 'to know'.

Old Irish. *rːo=fʸ'idʸ-ərʸ* ⟨*ro-fitir*⟩.

Proto-Brittonic. **gwıð-* is retained in all languages.

Proto-Germanic. Proto-Germanic knew a basic opposition between **wait-a-* 'to know (smth.; a situation), know that...' and **kann-a-* 'to know (smbd.), be acquainted with', although in some modern languages (e.g. English) this opposition became neutralized in favour of **kann-a-*.

Proto-Slavic. The opposition **vayd-eː-* ⟨**věděti*⟩ 'to know (a situation), know that...' / **znaː-* ⟨**znati*⟩ 'to know, be acquainted with an object/person' can be reconstructed for Proto-Slavic. It is retained in West lects, but tends to be simplified in favor of **znaː-* in the South and East. According to the GLD specification, we fill the slot with **vayd-eː-*.

Proto-East Baltic. The opposition *žin-aː-* 'to know (a situation), know that...' / *žin-* 'to know, be acquainted with an object/person' is retained in both Lithuanian and Latvian.

Proto-Iranian. The opposition *wid-* 'to know (a situation), know that...' / *ʒan-* 'to know how to, be acquainted with an object/person' is found in Avestan and can be reconstructed for Proto-Iranian (the same opposition is attested in Old Indian). In the majority of Iranian languages, *wid-* 'to know (a situation), know that...' in its primary meaning was superseded by *ʒan-*. In some languages, the meaning 'to know' is expressed with such verbs as *grab-* 'to grab / seize / take' etc.

Proto-Samoyed. *tänä-mä-* (Janhunen 1977: 157) is attested in most daughter languages. The word is derived from Proto-Samoyed *tänä-* 'to remember'.

Etymological notes. The lexical opposition between two verbs for 'to know' is characteristic for IE languages. It can be reconstructed at least for the Proto-Inner IE level as *woyd-* 'to know (a situation), know that...' / *gʸnoː-* (*gʸneh₃-) 'to know how to, be acquainted with an object/person'. In accordance with Kassian et al. 2010, we take the first verb for our wordlists. Hittite *sakk-* and Latin *sk-iː-* represent the not infrequent semantic shift 'to cut' > 'to know' (cf. the same for the Tocharian verb) being derived from IE *sek-* 'to cut' with different morphological patterns. Because of this we formally treat Hittite *sakk-* and Latin *sk-iː-* as independent formations and mark them with different indexes in the derivational drift-free dataset (Stage-2).

46. 'leaf'.

Old Irish. *dulʸnʸe ~ dulʸːe* ⟨duilne ~ duille⟩.

Proto-Brittonic. *dal-* is retained in all languages.

Proto-Germanic. The old stem *lawb-a-* 'leaf' still has singulative usage in Gothic and some West Germanic languages (e.g. English ⟨leaf⟩), but already in medieval times shows a strong tendency to be reserved for the collective plural term 'leafs, foliage', while the singulative 'leaf' is expressed by the innovation *blad-an*, formerly probably = 'blade, edge' (cf. the compound ⟨lauf-blað⟩ 'a single leaf' in Old Norse). We select *lawb-a-* as the more archaic term.

Proto-Slavic. *liːst-u* ⟨*listъ⟩ is retained in all three subgroups.

Proto-East Baltic. **lap-a- ~ *lap-a:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **war-ka-* is retained in Avestan and all subgroups. In some languages, mostly Eastern Iranian, polysemy 'feather / leaf' is attested. The further etymology is unclear.

Proto-Samoyed. **yapä* (Janhunen 1977: 41) is retained in Northern Samoyed, Kamass and Selkup.

Etymological notes. Tocharian **pəlt-a:* is a participle in **-t-* from IE **bʰlo:-* (**bʰleh₃-*) 'to bloom, flourish', whereas Ancient Greek *pʰúll-o-n* (*-ll-* < **-ly-*) and Latin *fol-i-um* represent another derivative from the same verb with the **-yo-*suffix; because of this we assign two different index to the Tocharian and Ancient Greek-Latin forms in the derivational drift-free dataset (Stage-2).

47. 'to lie'.

Old Irish. *l:aɣ ʸ-əðʸ ⟨laigid⟩*.

Proto-Brittonic. **gwor=weð-* is retained in all languages.

Proto-Germanic. **leg-yan-* is retained in all ancient and modern languages.

Proto-Slavic. **lež-a:- ⟨*ležati⟩* (< **leg-e:-*) is retained in all three subgroups.

Proto-East Baltic. **gul-e:-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **cay-* is retained in Avestan and has external cognates. Descendants of this root in later Iranian languages have different meaning: 'to sleep' (Wakhi), 'to be sick, ill' (Ossetic). In the majority of Iranian languages the term 'to lie' is derived from verbs with original meanings 'to sleep', 'to fall' etc.

Proto-Samoyed. **kɨy-tV-* (Helimski 1997: 280–281), going back to Proto-Uralic **kuyi-* 'to lie', is retained in Mator. The Uralic verb **kuyi-* 'to lie' is a likely cognate of Indo-European **kʸey-* 'to lie'.

Etymological notes. The opposition **kʸey-* 'to lie (stative)' vs. **legʰ-* 'to lie down (change-of-state)' is likely to have existed in Proto-IE. Predictably such an opposition is not very stable and in various groups these verbs competed and superseded each other. In the homoplasy-optimized dataset (Stage-3), we mark the reflexes of **legʰ-* with three different indexes

for Tocharian, Old Irish-Germanic and Slavic, assuming that the semantic expansion *legʰ-
'to lie down' > 'to lie / to lie down' took place in these groups in parallel. We prefer to keep
the same index for the Old Irish and Germanic reflexes of *legʰ- in the homoplasy-
optimized dataset (Stage-3), since from the formal point of view the meaning shift *legʰ- 'to
lie down' > 'to lie / to lie down' can be postulated already for the Proto-Italic-Germanic-
Celtic level (the virtual absence of the verb *kʸey- in Italic, Germanic and Celtic speaks in
favor of this scenario). See notes on 'to sit' for the similar opposition between the stative
and change-of-state verbs.

48. 'liver'.

Old Irish. *oːa* ⟨*óa*⟩.

Proto-Brittonic. *a(v)üː* is retained in all languages. The reconstruction is approximate;
Welsh has inexplicable variants ⟨*afu*⟩ (Southern dialects) and ⟨*iau*⟩ (Northern dialects).

Proto-Germanic. *libr-oː* is retained in all ancient and modern languages (although not at-
tested in Gothic).

Proto-Slavic. *entr-a* ⟨*ętro*⟩ is retained in the South and West subgroups. In East Slavic and
some West Slavic lects, superseded with either *peč-en-i* 'a roasted one' (← *pek-* 'to roast,
bake') or *antr-ab-aː* ⟨*ǫtroba*⟩ 'entrails, intestine'.

Proto-East Baltic. *yek-n-* is retained in both Lithuanian and Latvian. In modern Lithuanian
superseded with ⟨*kẽpenys ~ kẽpenos*⟩ — a loan-translation of the East Slavic term for 'liver'
based on *kep-* 'to roast, bake'.

Proto-Iranian. The heteroclitical paradigm *yak-ar* / *yak-n-*, having direct Indo-Aryan and
PIE cognates, does not survive in this shape in any of Iranian languages. The direct stem
*yak-ar ~ *yak-r-* is retained in Avestan and all Iranian subgroups. Descendants of the
oblique stem *yak-n-* can be found in Pashto and some other Eastern Iranian languages.

Proto-Samoyed. *mɨtə̆* (Janhunen 1977: 93–94), retained in all Samoyed languages save Ma-
tor, goes back to Proto-Uralic *mɨksa* 'liver'.

49. 'long'.

Old Irish. *foda ~ fada* ⟨*fota ~ fata*⟩.

Proto-Brittonic. **hiːr* is retained in all languages.

Proto-Germanic. **lang-a-* is retained in all ancient and modern languages.

Proto-Slavic. **dilg-* ⟨**dьlg-*⟩ is retained in all three subgroups.

Proto-East Baltic. **ilg-a-* means generic 'long' in Lithuanian, having narrowed its meaning to 'long (temporal)' in Latvian.

Proto-Iranian. **darg-a-* is retained in Avestan and all subgroups. It is superseded by **bṛʒ-* 'high' in many Eastern Iranian languages.

Proto-Samoyed. **yʋmpə̆* (Janhunen 1977: 37) is retained in all daughter languages save Nganasan.

Etymological notes. Despite some phonological discrepancies, the comparison between Hittite *talug-i-*, Ancient Greek *dolikʰ-o-* ⟨δολιχός⟩ (a basic term for 'long' in Homer), Albanian *ʒa-tə*, Latin *lˈong-us*, Germanic **lang-a-*, Slavic **dilg-*, Baltic **ilg-a-*, Old Indic *diːrgʰ-ˈa-*, Iranian **darg-a-* is straightforward and generally accepted by experts.

50. 'louse'.

Old Irish. *mʸiːl* ⟨*míl*⟩. Polysemy: 'animal / louse'. Lucht (2007: 247) gives an extremely rare *sar ~ sor* ⟨*sar ~ sor*⟩ 'louse' as the main unmarked Old Irish term on the ground that 'louse' is the only meaning of this word. In Modern Irish the main word for louse is ⟨*míol*⟩ 'animal; insect, creature; louse', while ⟨*sor*⟩ means 'animal louse, tick'. There is no reason to think that the situation in Old Irish was different.

Proto-Brittonic. **lːow* is retained in all languages.

Proto-Germanic. **luːs* is retained in all ancient and modern languages (but is not attested in Gothic).

Proto-Slavic. **vuš-i* ⟨**vъšь*⟩ is retained in all three subgroups.

Proto-East Baltic. **ut-i- ~ *ut-yaː* is retained in both Lithuanian and Latvian.

Proto-Iranian. *cwiš-* is retained in all subgroups and is very stable, although its descendants often have phonetic irregularities. Note that its reflex in Avestan, the hapax *spiš-*, means 'moth'.

Proto-Samoyed. *vnçu* (Janhunen 1977: 18) is attested in Forest Nenets, Enets, Nganasan, Kamass and Selkup, whereas *pănsV* (Helimski 1997: 246) is attested in Tundra Nenets and Mator with derivatives in Forest Nenets and Enets. It seems clear that both roots were present in Proto-Samoyed, but the semantic difference between them is not clear.

51. 'man'.

Old Irish. *fʲer* ⟨*fer*⟩. Polysemy: 'man (as opposed to woman) / husband'.

Proto-Brittonic. *gwur* is retained in Welsh and Cornish. Replaced in Breton as the main word for 'man' by ⟨*gwaz*⟩, related to Welsh ⟨*gwas*⟩ 'boy / servant'.

Proto-Germanic. It seems reasonable to set up a weak opposition between *wir-a-* 'man = (grown) male human being' and *mann-* 'man = human being (in general)' for Proto-Germanic, although semantic boundaries between the two items may have begun to be blurred already on the proto-level, so that the derivate *mann-iska-* was already used to expressly denote the general semantics of 'human being'. As for *wir-a-*, it is still well preserved in the meaning 'male human being' in Gothic and some ancient West Germanic languages (cf. the Old English opposition between ⟨*wer*⟩ 'man' and ⟨*wíf*⟩ 'woman'), but is quickly specialized in the meaning 'husband' (Scandinavian, where the default word for 'male human being' is now *karla-*) or lost altogether, merging with *mann-*.

Proto-Slavic. *manž-i* ⟨*mǫžь*⟩ or its suffixed derivatives are retained in all three subgroups.

Proto-East Baltic. *vi:r-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *nar-* is the main designation of 'man' in Avestan and is retained with this meaning in almost all subgroups. In many languages, the polysemy 'man / male' is attested, while some languages (Balochi, Parachi, Shughni etc.) preserve only the second meaning. The stem *vi:ra-* basically means 'man' as a hero in Avestan, together with Old Indic evidence (Vedic *wi:r-'a-* 'hero, a brave or eminent man') it should point to a Proto-Indo-Iranian semantics. The Avestan collocation ⟨*pasu vīra*⟩ 'animals and men' has the direct Umbrian cognate ⟨*uiro pequo*⟩ and thus belongs to *Dichtersprache*. In other Iranian lan-

guages, *viːra-* can have different meanings, e.g., 'man' as a hero in Middle Persian (a retention), 'husband' in Yaghnobi and Sogdian.

Proto-Samoyed. *tepä* (Janhunen 1977: 163) is retained in Mator, Kamass and Selkup. Nganasan has a derivative with the meaning 'boy'. In Northern Samoyed this word was replaced by Proto-Samoyed *kvə̆-sv* 'person'.

Etymological notes. Two terms are in criss-crossed configuration: (1) *wiː-ro-* (*wiH-ro-*) in Italic-Germanic-Celtic and Baltic, (2) *əner-* (*h₂ner-*) in Greek-Armenian and Indo-Iranian. Out of these, *əner-* (*h₂ner-*) seems more archaic due to its morphological primariness and better distribution. In its turn, *wiː-ro-* looks like a more recent introduction, since *wiː-ro-* 'man' is a substantivized adjective in *-ro-* from *wiː-* (*wiH-*) 'strength', the adjective status of *wiː-ro-* being retained in Tocharian. In light of this, we posit *əner-* (*h₂ner-*) as a Proto-Inner IE term for 'man' which was superseded with *wiː-ro-* (*wiH-ro-*) 'youthful and strong' > 'man' in the Italic-Germanic-Celtic clade and separately in Baltic (either independently or under the Germanic influence). Baltic *viːr-a-* is marked with a separate index in the homoplasy-optimized dataset (Stage-3).

52. 'many'.

Old Irish. *il* ⟨il⟩. Used both as an independent word and as a first part of compounds. Such compounds can be formed with practically any plural or collective noun. This phenomenon reflects an old word order "adjective — noun", replaced in Old Irish by the order "noun — adjective".

Proto-Brittonic. *lːower*, retained in Welsh ⟨llawer⟩ 'many, large number, much' and Cornish ⟨lower⟩ 'many, much; enough; sufficient', is cognate to Old Irish ⟨loor, lour⟩ 'enough, sufficient'. Cornish also has ⟨lyes⟩ 'many', related to Welsh ⟨lliaws⟩ 'multitude, abundance' and Breton ⟨lies⟩ 'numerous' < Proto-Brittonic *lːiːɔːs* < PIE *pleː-* (*pleh₁-*). It is unclear, which of these words, if any, was the main synonym for 'many' in Proto-Brittonic.

Proto-Germanic. The old adjective *manag-a-* is retained in all ancient and most modern languages, but sometimes is replaced by the old adverbial form *filu* 'very; much' (as in German ⟨viel⟩).

Proto-Slavic. *munag-a* ⟨*mъnogo⟩ is retained in all three subgroups.

Proto-East Baltic. **dawg-i* is retained in both Lithuanian and Latvian.

Proto-Iranian. **par-u-* is retained in Avestan and all Iranian subgroups, sometimes with a polysemy 'many / very'; it has direct Indo-Aryan cognates.

Proto-Samoyed. **oykkʋ ~ *oytkʋ ~ *oyçkʋ ~ *oyskʋ* (Janhunen 1977: 29) is retained in most daughter languages.

Etymological notes. Several involved forms represent an adjective in **-u-* from the verb **pleː-* (**pleh₁-*) 'to be full': Ancient Greek *pol-ǘ-*, Old Irish *il*, Old Indic *pur-ˈu-*, Iranian **par-u-*. A different morphological pattern is attested in Brittonic **lːiːɔːs*, which goes back to the same verbal root modified with the complex suffix **-yoːs-t-*. Because of this we mark the Brittonic form with a separate index in the derivational drift-free dataset (Stage-2). The match between Germanic **manag-a-* and Slavic **munag-a* ⟨**mъnogo*⟩ is clearly of areal origin. These forms together with the related Celtic stem **menekki-* 'frequent' demonstrate irregular sound correspondences and a non-standard phonotactic structure that could eventually imply a borrowing from a certain non-Indo-European substrate source. We mark the Germanic and Slavic form with two different indexes in the homoplasy-optimized dataset (Stage-3).

53. 'meat'.

Old Irish. *fʲeo̯lʲ* ⟨*feóil*⟩ 'flesh, both of living and dead bodies, often of meat: distinguished from cooked flesh'. Secondary synonym: ⟨*carna*⟩ 'flesh, meat' (< Latin). The old Indo-European word for 'meat' is preserved in a derivative ⟨*mír*⟩ 'bit, morcel' < **meːms-ro-*.

Proto-Brittonic. **kiːg*, cognate to Old Irish ⟨*cích*⟩ 'female breast', is retained in all languages.

Proto-Germanic. **memz-a-* is the best candidate for Proto-Germanic 'meat', despite being preserved only in one branch (Gothic); the external argument (of cognates in other IE branches) is defining here. Replaced in Proto-Scandinavian by **ketwa-* (> Old Norse ⟨*kjǫt*⟩, etc.⟩ of unclear origin; in Proto-West Germanic by **flaiska-*, also of unclear origin.

Proto-Slavic. **mens-o* ⟨**męso*⟩ is retained in all three subgroups.

Proto-East Baltic. **mens-aː* is retained as 'meat' in Lithuanian dialects (in Literary Lithuanian superseded with a cognate East Slavic loan), having narrowed its meaning to 'body; flesh' in Latvian.

Proto-Iranian. The stem **gaw-* (whose original meaning is 'cow / bull') or its derivative **gaw-šta-* are retained in Avestan and all subgroups.

Proto-Samoyed. **vyv* (Janhunen 1977: 17) preserves the sense 'meat / flesh' in Kamass and Selkup. In Northern Samoyed this word retains only the meaning 'flesh', being replaced in the sense 'meat' by **ŏm-sv* (originally 'food', derived from **ŏm-* 'to eat'). In Mator, the root **vyv* is also replaced by **ŏm-sv*, being preserved only in the derivate 'raw'.

54. 'moon'.

Old Irish. *eːske ~ eːsʲkʲe* ⟨*ésca(e) ~ éisce*⟩. This word may be a derivative of obsolete ⟨*éig*⟩ 'moon' (attested only in glossaries). Note also ⟨*lúan*⟩ (< PIE **lowk-sno-*) in ⟨*día lúain*⟩ 'Monday' — a second candidate for the status of the Proto-Irish term 'moon'.

Proto-Brittonic. **lːoɣr* is retained in all languages, but replaced in Modern Welsh as the main word for 'moon' by ⟨*lleuad*⟩, derived from ⟨*lleu*⟩ 'light'.

Proto-Germanic. **meːn-oːn* is retained in all three branches, although in Old Norse and Icelandic (but not Faroese!) mostly as an archaism, replaced by ⟨*tungl*⟩ of unclear origin.

Proto-Slavic. There are two candidates intertwined within the subgroups: (1) **meːs-enk-u* ⟨**měsękъ, *měsęcь*⟩, and (2) **loːn-aː* ⟨**luna*⟩. Out of them, **meːs-enk-u* has the advantage. First, it has a wider distribution among Slavic languages. Second, it can be observed in some cases of language evolution such as from Old Russian to Modern Russian that **meːs-enk-u* is gradually being superseded by **loːn-aː* in the meaning 'moon'. Third, the meaning 'month' is normally expressed with **meːs-enk-u* even in those lects where **loːn-aː* is a basic word for 'moon' (e.g., Modern Russian), which points out that **loːn-aː* is a later introduction. There is no doubt that both **meːs-enk-u* and **loːn-aː* are Proto-Slavic stems whose meaning was 'moon', but the available evidence suggests that it was **meːs-enk-u* which served as a default and neutral term for 'moon' in the protolanguage.

Proto-East Baltic. **meːn-oː*, obl. **meːn-es-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **maːh-* or its derivatives are retained in Avestan and all Iranian subgroups. It has direct Indo-Aryan and IE cognates.

Proto-Samoyed. **kɨy* (Janhunen 1977: 69), retained in Nganasan, Mator and Kamass, goes back to Proto-Uralic **kɨwi* 'moon'.

Etymological notes. Albanian *hən-ə* and Old Indian *čand-ra* represent different nominal formations from the verbal root **(s)kend-* 'to shine (*vel sim.*)' with the **-na:-* and **-ro*-suffix respectively; because of this we assign two separate indexes to the Albanian and Old Indian forms in the derivational drift-free dataset (Stage-2). A similar case is Armenian *lusin* and Latin *lu:n-a* which are derived from the verb **lewk-* 'to shine' with different suffixes: possibly **-eno-* in Armenian (Olsen 1999: 465) and **-sna:* in Latin; we mark the Armenian and Latin forms with two separate indexes in the derivational drift-free dataset (Stage-2).

55. 'mountain'.

Old Irish. *sʸlʸiạʋ* ⟨*slíab*⟩.

Proto-Brittonic. **monıð* is retained in all languages.

Proto-Germanic. **berg-a-* is well preserved as the default equivalent for 'mountain' in most West Germanic languages (except for English, where it has been replaced by the French borrowing ⟨*mountain*⟩) and in East Scandinavian languages (Swedish, Danish). West Scandinavian languages tend to replace it with reflexes of **felz-a-* 'rock' (Old Norse ⟨*fjall*⟩, etc.); Gothic ⟨*fairguni*⟩ < Proto-Germanic **fergunya-*, a somewhat rare term that more frequently applies to 'mountain ranges', 'mountainous terrain' in other ancient Germanic languages and may have been secondarily "singularized" in Gothic.

Proto-Slavic. **gar-a:* ⟨**gora*⟩ is retained in all three subgroups.

Proto-East Baltic. **kaln-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **gar-i-* is retained in this meaning in Avestan, Southwestern, Eastern languages and Ormuri. In Northwestern languages this root is preserved only in altered meanings (compare Kurdish ⟨*gir, girik*⟩ 'hill / height'). In many languages a semantic change 'mountain' > 'stone' is found.

Proto-Samoyed. **wʋrʋ*, retained in Enets, Nganasan and Kamass, goes back to Proto-Uralic **wara* 'mountain'.

Etymological notes. Latin *mo:n-s*, obl. *mont-* and Brittonic **monıð* represent separate derivatives from the root **mon-* with the suffixes **-ti-* and **-yo-* respectively. The starting root is likely to be a verbal one since, first, suffixal **-ti-* normally forms deverbatives and, second, the verbal status of **mon-/*men-* can be retained in the Latin stative *e:=min-e:-* 'to stick out,

protrude'. Because of this we prefer to mark the Latin and Brittonic forms with different indexes in the derivational drift-free dataset (Stage-2).

56. 'mouth'.

Old Irish. b^ye:l ⟨bél⟩. Polysemy: 'lip / mouth'. This is the most frequent word for 'mouth' in textual examples in DIL; modern Goidelic words for 'mouth' also continue this word. However, attestations in Old Irish glosses are generally in the plural and mean 'lips'. On the other hand, the meaning 'mouth' is frequently rendered in Old Irish glosses by another word, g^yin ⟨gin⟩. We provisionally list both words as synonyms. The Proto-Indo-European word *o:s- (*(H)oHs-) 'mouth' is preserved in Old Irish as an extremely rare a: ⟨á⟩ 'mouth'.

Proto-Brittonic. *gen-ow is retained in all languages, but the main word for 'mouth' in Modern Welsh is rather ⟨ceg⟩ (perhaps borrowed from Old English ⟨céce⟩ 'the swallow, fauces; mouth, cheek'). *gen-ow is the plural form of *gen, retained in Welsh and Middle Breton in the original meaning 'jaw, chin' (the word goes back to PIE *g^yenu- 'jaw, chin').

Proto-Germanic. *munθ-a- is retained in all ancient and modern languages.

Proto-Slavic. Dualia tantum *o:st-a: ⟨*usta⟩ is retained at least in the South and West subgroups and thus can be safely reconstructed as the Proto-Slavic term for 'mouth'. The morphological structure, however, suggests that the original meaning of *awst-a: (> *o:st-a:) in Pre-Proto-Balto-Slavic should have been 'lips' *vel sim.* (this conclusion is supported by the Old Indic *comparanda*).

Proto-East Baltic. *mut-ya: is retained as 'mouth' in Lithuanian dialects and Latvian. This stem superseded the Proto-Balto-Slavic term *awst-a: 'mouth', which is retained in Old Prussian.

Proto-Iranian. *a:h- is retained only in Avestan and Khotanese, but its antiquity is proved by direct Indo-Aryan and IE cognates.

Proto-Samoyed. *aŋ (Janhunen 1977: 20), retained in all Samoyed languages, goes back to Proto-Uralic *aŋi 'mouth / opening'.

57. 'name'.

Old Irish. *anʸmʸ ⟨ainmm⟩*.

Proto-Brittonic. **anw* is retained in all languages.

Proto-Germanic. **nam-oːn* is retained in all ancient and modern languages.

Proto-Slavic. **yiːm-en ⟨*jьmę⟩* is retained in all three subgroups. The accentual evidence points out that the first vowel is to be reconstructed as a long one.

Proto-East Baltic. **vard-a-* is attested as 'name' in Lithuanian and with polysemy 'word / name' in Latvian. Since external *comparanda* points to the original meaning 'word' for this stem, the polysemy 'word / name' is to be reconstructed for Proto-East Baltic. The polysemy was retained in Latvian, but got simplified in Lithuanian. The Proto-Baltic term **inmen-* 'name' is attested in Old Prussian (for another vowel grade see the Yotvingian form).

Proto-Iranian. **naːm-an-* is retained in Avestan and all subgroups. It is a very stable word with direct Indo-Aryan and IE cognates.

Proto-Samoyed. **nim* (Janhunen 1977: 102), retained in all Samoyed languages, goes back to Proto-Uralic **nimi* 'name', which is a likely cognate of Indo-European **nom-n̥* (**h₁nom-n̥*) 'name' (Illich-Svitych 1976: 82–83).

58. 'neck'.

Old Irish. *braːɣe ~ braːɣʸe ⟨brágae ~ bráige⟩*.

Proto-Brittonic. **guðug* is retained in Welsh and Breton.

Proto-Germanic. **xals-a-* is retained in almost all ancient and modern Germanic languages, except for English (where ⟨*neck*⟩ goes back to the original Proto-Germanic word for 'nape of the neck').

Proto-Slavic. **šiːɣ-a: ⟨*šija⟩* is retained in all three subgroups. Its derivation from the verb **šiː-* 'to sew' seems problematic both morphologically and semantically.

Proto-East Baltic. **kakl-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **griː-waː* is retained in Avestan, Southwestern and Eastern Iranian languages, its antiquity is supported by direct Indo-Aryan cognates. A widespread stem

*gard-a-na- ~ *gart-a-na- 'round' with the semantic development 'round' > 'neck' represents a later derivative from *gart- ~ *gard- 'to turn'.

Proto-Samoyed. *wayk-kšə (Janhunen 1977: 173) is retained in all languages save Selkup. Replaced in Selkup by reflexes of Proto-Samoyed *soy 'throat' (Janhunen 1977: 142). The word *wayk-kšə is etymologically a dual form of *wayk 'shoulder' (< Proto-Uralic *wolka 'shoulder'), itself preserved only in Selkup, being replaced in most other languages by *mǒrkä 'shoulder'.

59. 'new'.

Old Irish. *noːe ⟨nóe⟩. Polysemy: 'new, recent, fresh'.

Proto-Brittonic. *nowɪð is retained in all languages.

Proto-Germanic. *new-ya- is retained in all ancient and modern languages.

Proto-Slavic. *nav- ⟨*nov-⟩ is retained in all three subgroups.

Proto-East Baltic. *naw-ya- is retained in Lithuanian, having been superseded with *yawn-a- 'young' in Latvian.

Proto-Iranian. *naw-a- is retained in Avestan and in all subgroups. It is a very stable stem with direct Indo-Aryan and IE cognates.

Proto-Samoyed. *nʸʊrpV (Helimski 1997: 315) is retained in Enets and Mator. No other word for 'new' is reconstructible for Proto-Samoyed.

60. 'night'.

Old Irish. *aðəɣʸ ⟨adaig⟩. The old IE word for 'night' is retained only in some compounds and in the adverb ⟨innocht⟩ 'tonight'.

Proto-Brittonic. *nos is retained in all languages.

Proto-Germanic. *naxt-(i)- is retained in all ancient and modern languages.

Proto-Slavic. *nakt-i ⟨*noktъ⟩ is retained in all three subgroups.

Proto-East Baltic. **nakt-i-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **xšap-* is retained in Avestan and all Iranian subgroups; it has direct Indo-Aryan cognates. The stem **naxt-*, which also has reliable Indo-Aryan and IE etymology, is retained only in Avestan (adverbially) and Wakhi (noun, adverb). It is natural to suppose that Wakhi ⟨*naγ̌d*⟩ 'night / at night' represents a backward formation from some kind of adverb containing the root **naxt-*.

Proto-Samoyed. **pi* (Janhunen 1977: 123) is retained in all Samoyed languages.

Etymological notes. The match between Hittite *sp-ant-* (< **ksp-ant-*) and Iranian **xšap-* violates the tree topology. Although **ksep-* most likely is the Proto-IE term for 'night', there is little doubt that in Proto-Inner IE it was superseded with **nog^{wh}-t-* 'night' (whose original Proto-IE meaning was 'dusk, evening' as follows from the Anatolian data). Thus Iranian **xšap-* represents a backward semantic development, and we mark this form with a separate index in the homoplasy-optimized dataset (Stage-3).

61. 'nose'.

Old Irish. *sroːn* ⟨*srón*⟩. Polysemy: 'nostril / nose'. Cognate to Proto-Brittonic **froγn* > Welsh ⟨*ffroen*⟩, Breton ⟨*fron*⟩ 'nostril'.

Proto-Brittonic. **truγn* is retained in Welsh and Cornish. **friγ* (phonetic side of reconstruction unclear) is retained in Cornish and Breton. **friγ* (if reconstructed as **freγ*) may possibly have the same root as **froγn* 'nostril', but this is highly uncertain and we mark these two items as different roots.

Proto-Germanic. **nas-oː* is retained in most ancient and modern languages (though the word is not attested in Gothic); in a few Scandinavian languages, replaced with the innovation **nab-yan-* 'beak, bill, nib'.

Proto-Slavic. **nas-u* ⟨**nosъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **naːs-i-* is retained as 'nose' in Lithuanian and 'nostril' in Latvian.

Proto-Iranian. **nah-* / **naːh-* is retained in Avestan, Old Persian and all subgroups and has direct Indo-Aryan cognates. The stem **wayn-a-* ~ **wayn-i-*, derived from **wayn-* 'to see',

means 'nose' in some Iranian languages mostly of Southwestern and Northwestern subgroups.

Proto-Samoyed. *pɨyʋ ~ *puyʋ (Janhunen 1977: 122–123) is retained in all daughter languages save Nganasan.

62. 'not'.

Old Irish. *nʸiː* ⟨*ní*⟩.

Proto-Brittonic. *nɪ* is retained in all languages. Welsh and Breton developed a French-like double negation: Welsh ⟨*ni … ddim*⟩, Breton ⟨*ne … ket*⟩.

Proto-Germanic. *ne* is well represented in all three branches, although in Scandinavian languages it has been partially or completely replaced by an adverbial form (Old Norse ⟨*eigi, egi, ei*⟩, etc.) that probably began as an emphatic complementizer. In other languages the original morpheme is often augmented with various modifiers as well (English ⟨*not*⟩ < Old English ⟨*ne-á-wiht*⟩, etc.).

Proto-Slavic. *ne* ⟨*ne*⟩ is retained in all three subgroups.

Proto-East Baltic. *ne* is retained in both Lithuanian and Latvian.

Proto-Iranian. *na-* is retained in Avestan and all subgroups.

Proto-Samoyed. *e-* (Janhunen 1977: 26). Negative verb, inherited from Proto-Uralic and retained in all daughter languages.

63. 'one'.

Old Irish. *o̯in* ⟨*óen*⟩.

Proto-Brittonic. *üːn* is retained in all languages.

Proto-Germanic. *ayn-a-* is retained in all languages.

Proto-Slavic. *ediːn-* ⟨*edin-*⟩ is retained in all three subgroups. Morphological and etymological details are unclear. The traditional view is that medial *-iːn-* originates from *eyn-* 'one', but this analysis leaves the onset *ed-* without a convincing explanation.

Proto-East Baltic. *vien-a-* is retained in both Lithuanian and Latvian. The nature of initial *v-* is unclear, but the Lithuanian-Latvian forms can hardly be separated from Old Prussian *ayn-s* 'one'.

Proto-Iranian. *ay-wa-* is retained in Avestan and almost all other subgroups. Kümmel (2016) proposed to reconstruct Proto-Iranian *sya-* 'one' on the basis of the Saka and Ormuri numerals — a direct cognate of Hittite *sia-* 'one'. Whether or no one accepts Kümmel's analysis, the Saka and Ormuri data seems too scant to hypothesize that *sya-* might be a main numeral 'one' in Proto-Iranian which was further superseded with *ay-wa-* almost everywhere.

Proto-Samoyed. *o-p* (Janhunen 1977: 28) is retained in all daughter languages except Selkup, which has another derivative of the same root *o-*.

Etymological notes. Two sets of forms are in criss-crossed configuration on the Nuclear IE level. (1) *sem-* in Tocharian and Greek-Armenian. (2) A bulk of forms thought to contain a certain root *oy-* modified with various consonant extensions in other Inner IE groups. Inconsistency in phonetic and morphological details between the observed *oy-X-* forms could suggest that we are dealing with late introductions — either independent or contact-driven. Nevertheless, we prefer to refrain from final decision keeping this criss-crossing untouched in the homoplasy-optimized dataset (Stage-3). Moreover, it is not excluded that Hittite *si-* and Tocharian-Greek *sem-* are etymologically related and represent the single proto-morpheme (Hill 2012).

64. 'person'.

Old Irish. *dunʸe* ⟨duine⟩. Suppletive (sic!) plural *dọịnʸi* ⟨doíni⟩.

Proto-Brittonic. *dʉn* is retained in all languages.

Proto-Germanic. *mann-* or its derivative *mann-iska-* is retained in all three branches. See 'man' for additional comments on semantics.

Proto-Slavic. The morphologically unclear stem *čelaveːk-u* ⟨*čelověkъ⟩ is retained in all three subgroups.

Proto-East Baltic. *žmoː*, obl. *žmun-* is retained in Old Lithuanian as well as in Old Prussian. Cognate to *žem-y-aː* 'earth'.

Proto-Iranian. **mar-t-ya-* is retained in Avestan and all subgroups together with a closely related stem **mar-ta-* 'mortal / man'.

Proto-Samoyed. **kvə̆-sv* (Janhunen 1977: 61) retains the meaning 'person' in Mator and Kamass. In Northern Samoyed **kvə̆-sv* changed its meaning to 'man'. The word goes back to Proto-Uralic **kali-sʸa* 'person' (preserved also in Mansi), derived from Proto-Uralic **kali-* (> Samoyed **kvə̆-*) 'to die'.

Etymological notes. Latin *homoː*, obl. *homin-* and Baltic **žmoː*, obl. **žmun-* represent a derivative in **-on-* from the substantive stem **dʰgʸʰom-* 'earth'. The cognate Celtic stem **gdon-yo-* (> Old Irish *dunʸe*, Brittonic **dʉn*) goes back to the same root modified with another suffix. Since the meanings 'earth' and 'person' are semantically far from each other, we assign two separate indexes to the Latin-Baltic and Celtic forms in the derivational drift-free dataset (Stage-2). Nevertheless, after that two forms remain in criss-crossed configuration: (1) **dʰgʸʰ(o)m-on-* 'person' < 'earthling' in Latin and Baltic, (2) **mr̥-to- ~ *mor-to-* 'person' < 'mortal' in Armenian and Iranian. The second set has a better chance to be a parallel innovation since **m(o)r-to-* represents the most productive morphological pattern (*to*-participle from 'to die') and the most trivial semantic development. The important evidence for independent emergence of this term in Armenian and Iranian is that the stems in question are formed with different ablaut grades (zero in Armenian and *-o-* in Iranian). We mark the Armenian and Iranian forms with two separate indexes in the homoplasy-optimized dataset (Stage-3).

65. 'rain'.

Old Irish. *fʸlʸexʷəð* ⟨flechud⟩. Derived from *fʸlʸixʷ* ⟨fliuch⟩ 'wet'. Distinct from ⟨bráen⟩ 'rain, moisture, drop(s)' (mostly denoting light rain or drizzle) and ⟨folc⟩ 'heavy rain, wet weather' (Mac Mathúna 1978).

Proto-Brittonic. **glaw* is retained in all languages.

Proto-Germanic. **regn-a-* is retained in all ancient and modern languages.

Proto-Slavic. **dusd-y-i ~ *dusg-y-i* ⟨**dъždžь*⟩ is retained in all three subgroups.

Proto-East Baltic. **ley-tu-* is retained in both Lithuanian and Latvian. Derived from **ley-* 'to pour'.

Proto-Iranian. **waːr-a-* is retained in Avestan and all subgroups. It has Vedic cognates with the meaning 'water'.

Proto-Samoyed. **sʋr-ö* (Janhunen 1977 135–136), retained in Northern Samoyed and Mator, is derived from the verb **sʋrʋ-* 'to rain' (< Proto-Uralic **sʸaða-* 'to rain'). Kamass and Selkup have another derivative from the same verb.

Etymological notes. Tocharian B *swes-e* and Albanian *ši* represent different nominal formations from the verbal root **sew-* 'to rain (*vel sim.*)' (although morphological details remain unclear); because of this we assign two separate indexes to the Tocharian and Albanian forms in the derivational drift-free dataset (Stage-2).

66. 'red'.

Old Irish. *dʸerg* ⟨*derg*⟩ 'red, ruddy (used of color of blood, flame; also of orange or tawny hue as of ale, gold, etc.)'. Distinct from *r:uặð* ⟨*rúad*⟩ 'red, of a brownish or dark red (opp. to ⟨*derg*⟩ = bright red), oft. of blood-stains (opp. to ⟨*fland*⟩, the color of freshly shed blood)'. Secondary synonym: *flan:* ⟨*flann*⟩ 'red (esp. blood-red)'.

Proto-Brittonic. **r:ü:ð* is retained in all languages, although in Welsh the main synonym for 'red' is ⟨*coch*⟩, borrowed from Latin ⟨*coccum*⟩ 'the insect *Coccus ilicis* or the scarlet dye obtained from it'.

Proto-Germanic. **rawd-a-* is retained in all ancient and modern languages.

Proto-Slavic. **čirm-in-* ~ **čirv-y-en-* ⟨**čьrmьn-*, **čьrvl'en-*⟩. Both stems contains the root **čirv-* ~ **čirm-* 'worm' q.v. The pure adjective **čirm-in-* with the meaning 'red' is mostly attested in ancient languages (such as Old Church Slavonic and Old Russian), where it coexists with **čirv-y-en-*. The antiquity of **čirm-in-* is supported by the fact that the root variant in *-m-* is etymologically primary, see notes on 'worm'. Either already in Proto-Slavic or somewhat later via interdialectal contacts, **čirm-in-* ⟨**čьrmьn-*⟩ 'red' was being superseded with **čirv-y-en-* ⟨**čьrvl'en-*⟩, a regular passive participle from the factitive **čirv-i:-* 'to make red', literally 'to make worm-like' (derived from **čirv-i* 'worm').

Proto-East Baltic. **rawd-a-* is apparently attested as generic 'red' in some Latvian dialects. Its antiquity is proven by the meaning 'red' in Yotvingian. In Lithuanian, 'red' is expressed

71

with the derivative stem *rawd-a:-na-, whereas *rawd-a- has shifted to the meaning 'reddish'.

Proto-Iranian. *rawd-i-ta- is retained in Avestan and has direct Indo-Aryan and IE cognates. Khotanese ⟨rrusta-⟩ 'red' can be a cognate, if not from Khotanese ⟨rrus-⟩ 'to shine'. The stem *cuk-ra- ~ *cux-ra-, derived from the root *cawč- 'to shine', is widespread in Iranian languages with the meaning 'red', but its original semantics should likely be 'shining, bright; red (of fire)'.

Proto-Samoyed. *nʸar- (Janhunen 1977: 107–108) is retained in Northern Samoyed, Mator and Selkup.

Etymological notes. Although Rix, Kümmel, et al. (2001: 508) reconstruct primary verbal root *(h₁)rewdʰ- 'to make red', the only verbal form that can be safely projected onto the PIE level is stative *rudʰ-e:- (*rudʰ-eh₁-), derived from the adjectival root according to the rules of the Caland system. Therefore, Tocharian, Greek and Latin forms, going back to *rudʰ-ro- and Indo-Iranian forms, derived from another Caland variant, *rudʰ-i-, can be safely marked with the same index. Baltic, Germanic and Brittonic forms, going back to *rowdʰ-o-, can formally be viewed as deverbal. Nevertheless, although suffix *-o- itself does not participate in the Caland system, there are parallel cases when adjectives in *-o- are formed from adjectival roots participating in the Caland system in the absence of any verbal correlate (e.g., *nogʷ-o- 'naked' alongside the Caland form *negʷ-mo-). Because of this we mark with the same index all adjectival forms containing the root *rewdʰ- in the derivational drift-free dataset (Stage-2).

67. 'road'.

Old Irish. sʲlʲiɣe ⟨slige⟩ is perhaps the main word for 'road' in Old Irish. ⟨sét⟩, preferred by Lucht (2007: 318), rather means 'way'.

Proto-Brittonic. *hɪnt, cognate to Old Irish ⟨sét⟩ 'way', is retained in Breton. North Welsh and Cornish words for 'road' are borrowings: North Welsh ⟨ffordd⟩, Cornish ⟨ford⟩ < Old English ⟨ford⟩ 'ford'. South Welsh has unetymologized ⟨heol⟩ 'road'.

Proto-Germanic. *weg-a- is the most common root to denote any general path of pedestrian or transport communication in most Germanic languages. Replacements such as Eng-

lish ⟨road⟩ < 'riding' are quite rare (although numerous synonyms exist denoting specific types of roads and paths in most languages).

Proto-Slavic. *pant-i ⟨*pǫtь⟩ is retained in all three subgroups; it has reliable external cognates with the same meaning. It competes with *darg-a: ⟨*dorga⟩ which is also widely used for 'road' in Slavic. The etymology of *darg-a: (deverbative from *dirg-a:- ⟨*dьrgati⟩ 'to pull, tug') as well as its meanings such as 'ravine, gully', 'uncultivated plot, pasture', 'footprint', attested at least in the West and South subgroups, suggests that in Proto-Slavic *darg-a: meant specifically 'track, rough path beaten by use' rather than generic 'road'.

Proto-East Baltic. *kel-i- is retained in both Lithuanian and Latvian.

Proto-Iranian. *pant-a:- ~ *paθ- is retained as a basic term in Avestan and all subgroups (however in middle and modern Southwestern and Northwestern languages it means specifically 'path / track / direction / advice, counsel'). Cf. the stem *ra:θ-i(:)-, *ra:θ-ya-, which sometimes acquires the semantics 'road' in Iranian languages, being derived from the word *raθa- 'chariot'.

Proto-Samoyed. *ätʋ (Janhunen 1977: 24) is retained in Mator, Kamass and Selkup. The root also has derivatives in Nenets.

68. 'root'.

Old Irish. frʲeːn ~ frʲeːm ⟨frén ~ frém⟩.

Proto-Brittonic. *gwreyð is retained in all languages.

Proto-Germanic. *wurt-i- is retained in almost all the ancient and modern languages, although English ⟨root⟩ is a Scandinavian reborrowing (the archaic form ⟨wort⟩ is no longer in common usage).

Proto-Slavic. *kar-en-i ⟨*korenь⟩ is retained in all three subgroups.

Proto-East Baltic. *šak-n-i- is retained in both Lithuanian and Latvian. Derived from *šak-a: 'branch, fork'.

Proto-Iranian. Both *wayx-a- and *rayš-a-ka- are retained in all subgroups (sometimes they are synonymous within one language). Neither stem has any certain external cognates. We treat them as synonyms.

Proto-Samoyed. *wʊnço* (Janhunen 1977: 171), retained in all Samoyed languages, goes back to Proto-Uralic *wançaw* 'root'.

Etymological notes. Phonetic correspondences between Ancient Greek *r̥ízd-a:*, Albanian *rəɲ-ə*, Latin *rˈa:d-i:k-s*, Celtic **wrid-a:*, Germanic **wurt-i-* are not fully regular, but we prefer to follow Vine 1999 in treating all these forms as cognates, because all of them share the shape *wrd* which can hardly be a chance coincidence. The likely PIE reconstruction is **wr̥:d-* (**wr̥h₂d-*).

69. 'round'.

Old Irish. *krunʸdʸ* ⟨*cruind ~ cruinn*⟩ 'round, globular, circular; compact, precise'.

Proto-Brittonic. **krun: ~ *krʉn:* is retained in all languages.

Proto-Germanic. Not properly reconstructible. Attested forms are either narrowly distributed and semantically/morphologically derived (Old Norse ⟨*kring-lóttur*⟩ < **kreng-* 'ring'; Old English ⟨*seonu-wealt, sine-wealt*⟩ < **welt-an-* 'to roll') or straightforwardly borrowed from Old French *ront* (all West Germanic and East Scandinavian forms).

Proto-Slavic. **krang-l-* ⟨**krǫgl-*⟩ is retained in all three subgroups with the meaning 'round 3D / round 2D'. Derived from the substantive **krang-u* ⟨**krǫgъ*⟩ 'circle'.

Proto-East Baltic. **ap=val-u-* is retained in both Lithuanian and Latvian with the meaning 'round 3D / round 2D'. Prefixed derivation from **vel-* 'to felt, roll'.

Proto-Iranian. Proto-Iranian **gart-a-na-* is derived from **gart-* 'to turn'. It is retained in all subgroups. Avestan ⟨*zgərəsna-*⟩ 'round' represents another derivative from the same verb.

Proto-Samoyed. Not reconstructible for Proto-Samoyed.

70. 'sand'.

Ancient Attic Greek. We follow Beekes 2010: 89 and treat *ámm-o-s* ⟨ἄμμος⟩ as a substrate loanword.

Old Irish. *ganʸəõ* ⟨*gainem*⟩ 'sand, gravel'. Alternative candidate is *gʸrʸian* ⟨*grïan*⟩ 'gravel, sand, sea or river bottom'. We choose ⟨*gainem*⟩ since it is attested in the meaning 'sand' in

Würzburg glosses; in Modern Irish and Scottish Gaelic reflexes of ⟨*gainem*⟩ mean 'sand', while reflexes of ⟨*grïan*⟩ mean 'gravel, grit'.

Proto-Brittonic. **tɪwod* is retained in Welsh. Breton and Cornish words for 'sand' go back to **traeθ*, whose Welsh reflex means 'beach, (sea)shore, strand, coast'. This word is apparently borrowed from Latin ⟨*tractus*⟩ 'tract, region'.

Proto-Germanic. **sand-a-* is the common and stable equivalent for all Scandinavian and West Germanic languages. In Gothic, the equivalent is ⟨*mal-ma*⟩, corresponding to words meaning 'ore; malm; dry earth / sand' in other languages. Technically, the root **mal-* is eligible for inclusion, but **sand-a-* has better external comparisons and more promising distribution within the group (the merger of the meanings 'ore / malm' and 'sand' in Gothic is quite plausible).

Proto-Slavic. **peːs-uk-u* ⟨**pěsъkъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **smeːl-i-* or suffixed **smil-t-i-* are retained in both Lithuanian and Latvian.

Old Indic. We follow Lubotsky 2001: 302, 312 and treat ⟨*síkat-ā*⟩ together with its Iranian counterparts (virtual PIr. **cikat-aː*) as loans from an unknown source.

Proto-Iranian. A very unstable term. One of the possible candidates is **rayka-*, attested in Southwestern and Northwestern Iranian languages.

Proto-Samoyed. **yʌ̆* (Janhunen 1977: 36–37), the root for 'earth (soil)', also means 'sand' in Enets and Nganasan; its derivative **yʌ̆-rʋ* 'sandy / sandy bank' is reflected in all Samoyed languages. Another word for 'sand', **puʌ̆rʋ* (Helimski 1997: 251), attested only in Mator and Kamass, is more likely an areal isogloss than retention from Proto-Samoyed.

71. 'to say'.

Old Irish. *as=bʸʲerʸ* ⟨*as-beir*⟩.

Proto-Brittonic. **du=wed-* is retained in Welsh. Cornish and Breton have denominative verbs derived from **lːavar* 'speech'.

Proto-Germanic. Two main verbs of speech could be eligible for the slot 'to say' for Proto-Germanic. Of these, **sag-yan-* is the most common in all modern Germanic languages;

75

however, it is not attested in Gothic, and in such ancient languages as Old Norse and Old English it is able to mean both 'say' (introducing direct speech or an object clause) and 'tell, narrate, recount', the latter meaning seemingly more frequent in Old English (cf. also the Old Norse derivate ⟨*saga*⟩ 'tale, etc.'). On the other hand, the verb **kwiθ-an-*, while rapidly going out of usage in medieval and modern Germanic languages, is well attested in the meaning 'to say', and only this meaning, in all the ancient languages of all three branches, including Gothic. It seems logical to reconstruct the original opposition as **kwiθ-an-* 'to say' vs. **sag-an-* 'to tell'.

Proto-Slavic. An unstable item. The best candidate seems **rek-* ⟨**rekti*⟩, because this verb is used for 'to say' in ancient languages: Old Church Slavonic, Old Russian, Old Novgorod, Old Czech. It was superseded with **kaːz-aː-* ⟨**kazati*⟩ in some modern South lects and independently with prefixed **su=kaːz-aː-* ⟨**sъkazati*⟩ in all modern East lects.

Proto-East Baltic. **sak-iː-* is retained in both Lithuanian and Latvian. Derived from **sek-* 'to tell'.

Proto-Iranian. Suppletive roots **mraw-* and **wač-* are the main terms with the meaning 'to say' in Avestan; the first one is used in present, the second one — mostly in aorist and perfect. Old Indian data confirm that Avestan retains the original suppletive paradigm.

Proto-Samoyed. **mʋ-* ~ **mʋn-* (Janhunen 1977: 88), retained in Northern Samoyed and Kamass, goes back to Proto-Uralic **moni-* 'to say' (reflected in Mari and Hungarian).

72. 'to see'.

Old Irish. *að=kʸ'i:* ⟨*ad-cí*⟩. Suppletive perfect *að=k'on=dərʸkʸ* ⟨*ad-condairc*⟩. While ⟨*ad-cí*⟩ goes back to PIE root **kʷeys-*, perfect stem ⟨*ad-condairc*⟩ continues PIE **derkʸ-*.

Proto-Brittonic. **gwel-* is retained in all languages.

Proto-Germanic. **sexw-an-* is retained in all ancient and modern languages.

Proto-Slavic. **viːd-eː-* ⟨**viděti*⟩ is retained in all three subgroups.

Proto-East Baltic. **reg-eː-* is attested in Lithuanian dialects and Latvian. The original Proto-Baltic root **veyd-* 'to see' is only retained in Old Prussian.

Proto-Iranian. **wayn-* is retained in Avestan and all subgroups. **darc-* is used as perfect and aorist stem in Avestan and as past stem in Parachi and some other Eastern Iranian languages. **darc-* has reliable Indo-Aryan cognates. The third root **daːy-* usually forms past stem in all subgroups, but its Indo-Aryan cognates demonstrate another meaning.

Proto-Samoyed. **mănçV-* (Janhunen 1977: 86–87) is retained in Nenets, Enets, Kamass and Selkup.

Etymological notes. Three verbal roots are in criss-crossed configuration. (1) **weyd-* in Ancient Greek (aorist), Latin, Slavic. (2) **derkʸ-* in Old Irish (perfect), Indo-Iranian (perfect, aorist). (3) **sekʷ-* in Albanian, Germanic. Out of these, the Albanian-Germanic match is a transparent case of homoplasy since the Proto-Inner IE meaning of **sekʷ-* can be safely reconstructed as 'to follow'. We mark the Albanian and Germanic forms with two different indexes in the homoplasy-optimized dataset (Stage-3). The distribution and status of **weyd-* and **derkʸ-* are less unambiguous. Both look archaic and both have a good chance to have represented a suppletive aorist and/or perfect stem in the Proto-Inner IE paradigm 'to see'. In the absence of additional evidence we are forced to keep **weyd-* and **derkʸ-* untouched in the homoplasy-optimized dataset (Stage-3). It also remains intriguing what root was used for the present stem in that suppletive proto-paradigm in Proto-Indo-European.

73. 'seed'.

Old Irish. *sʸiːl* ⟨síl⟩.

Proto-Brittonic. **had* is retained in all languages.

Proto-Germanic. A complicated situation. The reconstructible nominal stem **frayw-a-*, of unclear origin, is well attested in Gothic and Scandinavian, but not in Western Germanic. On the other hand, all three branches also preserve reflexes of Proto-Germanic **seː-di-* 'seed', derived from **seː-* 'to sow': in West Germanic, it is the main equivalent for 'seed' in general (although in this branch it also competes with an alternate, and possibly even more archaic, derivate **seː-moːn*); in Scandinavian, its usage is more restricted (usually 'seed' = 'grain for sowing'), and in Gothic, it is only attested figuratively in the compound ⟨mana-sēþ-s⟩ 'mankind'. Internal logic would recommend reconstructing **frayw-a-* and dismissing the rest as products of secondary derivation; external argumentation, however, shows that, while there are no IE parallels to **frayw-a-*, morphological derivates from **seː-*

are freely used to form the main word for 'seed' in numerous other branches. Without attempting to resolve this controversy, we count both words as technical synonyms.

Proto-Slavic. **seː-m-en* ⟨**sěmę*⟩ is retained in all three subgroups. Derived from **seː-ye-* ⟨**sěti*⟩ 'to sow'.

Proto-East Baltic. **seː-tl-aː* is retained in both Lithuanian and Latvian. Derived from **seː-* 'to sow'.

Proto-Iranian. **tawx-man-* is retained in Avestan and all subgroups; this stable term has good Indo-Aryan cognates with the meaning 'progeny, offspring'.

Proto-Samoyed. Not reconstructible reliably. Nenets and Enets use reflexes of **sǎymä* 'eye' (Janhunen 1977: 132); in most other languages the word for 'seed' is not attested.

Etymological notes. The bulk of the forms involved consists of different nominal derivatives from the verb **seː-* (**seh₁-*) 'to sow'. (1) The **-men*-suffix in Latin *seː-men*, Slavic **seː-men*. (2) The **-lo*-suffix in Old Irish *sʸiːl*. (3) The **-to*-suffix from the zero-grade of the root in Brittonic **had*. (4) The **-ti*-suffix from the *e*-grade of the root in Germanic **seː-di-*. (5) The **-tlaː*-suffix in Baltic **seː-tl-aː*. Because of this we assign five separate indexes to these groups in the derivational drift-free dataset (Stage-2). A similar case is Ancient Greek *spérma* and Albanian *far-ə* (if inherited) which represent different formations from the verbal root **sper-/*spor-* 'to scatter, strew': the **-mn*-suffix in Greek and the thematic stem **spor-aː* in Albanian; because of this we mark the Ancient Greek and Albanian forms with two separate indexes in the derivational drift-free dataset (Stage-2). Note also that morphologically identical Latin *seː-men* and Slavic **seː-men* most likely represent parallel new formations as well and thus should have been disjoined in the homoplasy-optimized dataset (Stage-3), but we have no formal evidence for such a decision.

74. 'to sit'.

Albanian. The generic verb *ri* ⟨*rri*⟩ 'to stay' > 'to sit / to stand' formally has the suppletive aorist *ndɛy-* ⟨*ndej-*⟩. Actually, however, ⟨*ndej-*⟩ is very rarely used. Default expressions for the stative preterite meaning 'he was sitting/standing' are regular perfect forms from ⟨*rri*⟩.

Old Irish. *saðʸ-əðʸ* ⟨*saidid*⟩.

Proto-Brittonic. Welsh ⟨*eistedd*⟩ on the one hand, and Cornish ⟨*esedh-*⟩, Breton ⟨*azez-*⟩ on the other, are derived from the root **seð-* with different prefixes.

Proto-Germanic. **set-yan-* is retained in all ancient and modern languages.

Proto-Slavic. **seːd-eː-* ⟨**sěděti*⟩ is retained in all three subgroups with the meaning 'to sit'. Derived from **seːd-* ⟨**sěsti*⟩ 'to sit down'.

Proto-East Baltic. **seːd-eː-* is retained in both Lithuanian and Latvian. Derived from **seːd-* 'to sit down'.

Proto-Iranian. According to Avestan and external Indo-Aryan data, Proto-Iranian **aːh-* meant 'to sit' and **had-* meant 'to sit down'. **aːh-* is retained for the stative meaning 'to sit' in Avestan and Middle Iranian languages. In the majority of Iranian languages belonging to all subgroups **had-* acquired the meaning 'to sit', having superseded **aːh-*.

Proto-Samoyed. **ʋmtə̆-* (Janhunen 1977: 17–18) is retained in all Samoyed languages.

Etymological notes. The opposition **eːs-* (**h₁eh₁s-*) 'to sit, be sitting (stative)' vs. **sed-* 'to sit down (change-of-state)' is likely to have been present in Proto-IE. As in the similar case of **kʸey-* 'to lie' vs. **legʰ-* 'to lie down', it is natural that such an opposition was not very stable having tended to simplify in favor of one of the verbs in individual subgroups. In the homoplasy-optimized dataset (Stage-3), we mark the reflexes of **sed-* with three different indexes for Armenian, Italic-Germanic-Celtic and Balto-Slavic, assuming that the semantic expansion **sed-* 'to sit down' > 'to sit / to sit down' took place in these groups in parallel. We prefer to keep the same index for the Latin, Germanic and Celtic reflexes of **sed-* and the same index for its reflexes in Baltic and Slavic in the homoplasy-optimized dataset (Stage-3), since from the formal point of view the meaning shift **sed-* 'to sit down' > 'to sit / to sit down' can be postulated already for the Proto-Italic-Germanic-Celtic level and independently(?) for the Proto-Balto-Slavic level (the virtual absence of the verb **eːs-* in the aforementioned groups speaks in favor of this scenario).

75. 'skin (human)'.

Old Irish. *kʸnʸes* ⟨*cnes*⟩ 'skin (of body): surface; *in extended application* body, flesh; bosom, breast'. Alternative candidate is *krokʸən:* ⟨*croicenn*⟩ 'hide, skin'. We choose ⟨*cnes*⟩ because this word is primarily applied to human skin.

Proto-Brittonic. **kroen ~ *kroxen*. The first variant, going back to Proto-Celtic **krokno-*, is retained in Welsh, the second (< PC **krokkeno-*) — in Cornish and Breton. The word is cognate to Old Irish *krokʸən:* ⟨*croicenn*⟩ 'hide, skin', but lacks an Indo-European etymology.

Proto-Germanic. An unstable etymon. The most widespread candidate is **xu:d-i-*, sometimes narrowed down to 'animal skin' (English), but more often applicable to both animals and people.

Proto-Slavic. **kaz-y-a:* ⟨**koža*⟩ is retained in all three subgroups, applicable to humans and animals. Derived from **kaz-a:* ⟨**koza*⟩ 'she-goat'.

Proto-East Baltic. **a:d-a:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **pawasta-* can probably be analyzed as the past participle **pa=was-ta-* from the verbal root **was-* 'to wear clothes, be dressed' with prefix **(a)pa-*. The first meaning of this form is 'skin'; polysemy 'skin / bark' is found in many Iranian languages and can be reconstructed for the proto-language. The stem **pawasta-* is retained in Avestan in the meaning 'skin (especially on the head of a human)' and in Southwestern (Middle Persian ⟨*pōst*⟩ 'skin / hide', Modern Persian ⟨*pust*⟩ 'skin / hide / bark / peel / (nut)shell') and several Eastern Iranian languages (Wakhi ⟨*pist*⟩ 'skin / hide / bark', Sanglechi ⟨*pask*⟩ 'id.').

Proto-Samoyed. **kopʋ* (Janhunen 1977: 73), retained in all Samoyed languages, goes back to Proto-Uralic **kopa* 'skin'.

76. 'to sleep'.

Old Irish. *kon=t'ulʸi* ⟨*con-tuili*⟩.

Proto-Brittonic. **kusk-* is retained in all languages.

Proto-Germanic. External evidence shows that **swef-an-*, only preserved as the neutral equivalent for 'sleep' in the Scandinavian branch, is more archaic in this meaning than the alternate verb **sle:p-an-*, even though the latter has the neutral meaning 'sleep' in two out of three branches (Gothic and West Germanic). It should be noted that in Old English, both verbs still compete for the same meaning, implying that even in Proto-West Germanic, **sle:p-an-* (originally = 'to be weak, numb') may only have initiated the process of displacing **swef-an-* from its former role.

Proto-Slavic. **sup-aː-* ⟨**sъpati*⟩ is retained in all three subgroups.

Proto-East Baltic. **meyg-aː-* is attested as 'to sleep' in Lithuanian and 'to doze' in Latvian (generic 'to sleep' is expressed with 'to lie' in Latvian). Derived from Proto-Baltic **meyg-* 'to sleep', retained in Old Prussian. Also cf. Proto-Baltic **mig-* 'to fall asleep'.

Proto-Iranian. **hwap-* is retained in Avestan and all subgroups; it has direct Indo-Aryan cognates. Cf. also Avestan *hah-* 'to sleep' — a marginal and rarely used verb as compared with basic Avestan *hʷap-* 'to sleep'. The same concerns the opposition *sas-* vs. *swap-* in Vedic. Thus there is no evidence that **sas-* might be a basic root for 'to sleep' in Proto-Indo-Iranian or Proto-Iranian.

Proto-Samoyed. **kont-ö-* (Janhunen 1977: 73) is retained in all daughter languages. Similarity to Proto-Yukaghir **qont-oː-* 'to lie' may be due to borrowing in either direction.

77. 'small'.

Old Irish. *bʲeg* ⟨*becc*⟩.

Proto-Brittonic. **bɪxan* is retained in all languages, but in Welsh the main root for 'small' is phonetically similar ⟨*bach*⟩.

Proto-Germanic. Two stems, **smal-a-* and **luːtil-a-*, are reconstructible as synonyms for this meaning in Proto-Germanic, much as they are still preserved in modern English or Icelandic.

Proto-Slavic. **maːl-* ⟨**mal-*⟩ or its derivatives with diminutive suffixes are retained in all three subgroups.

Proto-East Baltic. **maž-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **kac-u-* is retained in Avestan and all subgroups. In modern languages usually superseded by other roots.

Proto-Samoyed. **üçä* (Janhunen 1977: 31) is retained in Northern Samoyed and Selkup.

Etymological notes. Following Kroonen (2013: 456) we regard Proto-Germanic **smal-a-* as unrelated to Proto-Slavic **maːl-*.

78. 'smoke'.

Old Irish. d^ye: ⟨*dé*⟩. Genitive singular $d^yiað$ ⟨*diad*⟩ with hiatus. The word goes back to post-PIE **d^hi:yot-* < **d^hu:yot-* (**u:y* > **i:y* by Thurneysen's rule) from the PIE root **d^huh₂-*.

Proto-Brittonic. *mug is retained in all languages.

Proto-Germanic. **rawk-i-* is retained everywhere except for English.

Proto-Slavic. **du:-m-u* ⟨*dymъ⟩ is retained in all three subgroups. Historically derived from the verb **do:-* ⟨*duti⟩ 'to blow'.

Proto-East Baltic. **du:-m-a-* is retained in both Lithuanian and Latvian.

Proto-Iranian. The stem **du:-ta-* is retained in all subgroups. **du:-ma-* is attested in Khotanese. No expression for 'smoke' in Avesta. Both **du:-ta-* and **du:-ma-* are regularly derived from the verbal root **daw-* / **du:-* 'to smoke' with direct Indo-Aryan and IE parallels.

Proto-Samoyed. *küntằ (Janhunen 1977: 79), the main word for 'smoke' in Enets, Nganasan and Mator, goes back to Proto-Uralic *künti, whose Finno-Ugric reflexes mean 'fog'. Cf. *kằčku (Janhunen 1977: 40) with reflexes meaning 'fog' (Enets, Nganasan) and 'smoke' (Selkup).

Etymological notes. The bulk of the involved forms consists of several nominal derivatives from the verb **d^hwex-* (**d^hweh₂-*) 'to smoke'. (1) The **-wi*-suffix in Hittite *tuxxu-i-*. (2) The **-mo*-suffix in Latin *fu:m-us*, Balto-Slavic **du:-m-a-*, Old Indic *d^hu:-ma-*. (3) The complex **-y-Vt*-suffix in Old Irish d^ye:. (4) The **-to*-suffix in Iranian **du:-ta-*. We assign four different indexes to these groups in the derivational drift-free dataset (Stage-2).

79. 'to stand'.

Albanian. See notes on 'to sit' for the suppletive aorist *ndɛy-* ⟨*ndej-*⟩.

Old Irish. *ses:ə̃* ⟨*sessam*⟩ 'standing' — verbal noun, used in constructions of the type 'be in one's standing' = 'to stand'. Alternative candidate: *do=ˈar^y=əs^y:ə-ðər* ⟨*do-airissedar*⟩ 'stands, stays, remains (oft. used to gloss Lat. *stare* and compds.)'. Both forms have the root *siss-*, going back to PIE reduplicated stem **sti-st-* (**sti-sth₂-*).

Proto-Brittonic. *sav-* is retained in all languages. This verb is derived from the noun *sav* (> Welsh ⟨*saf*⟩ 'a stand(ing), station, standpoint', Breton ⟨*sav*⟩ 'upright position') that ultimately goes back to PIE root *sta:-* (*steh₂*-).

Proto-Germanic. *stand-an-* is retained in all ancient and modern languages.

Proto-Slavic. *sta-ya:-* ⟨*stojati*⟩ is retained in all three subgroups, meaning 'to stand'. Derived from the verb *sta:-* ⟨*stati*⟩ 'to stand up'.

Proto-East Baltic. *sta:-v-e:-* is retained in both Lithuanian and Latvian. Derived from *sta:-* 'to stand up'.

Proto-Iranian. *sta:-* is retained in Avestan and all subgroups; this verbal stem has direct Indo-Aryan and IE cognates.

Proto-Samoyed. *nu-* ~ *nɨ-* (Janhunen 1977: 104) is retained in Nenets, Enets, Kamass and Selkup.

Etymological notes. The common Inner IE verb is *sta:-* (*steh₂*-) as in Ancient Greek *hí=ste:-mi*, Latin *sta:-re*, Slavic *sta-ya:-*, Baltic *sta:-v-e:-*, Old Indic *sthа:-*, Iranian *sta:-*, and finally Germanic *stand-an-* (via the chain of secondary formations within the verbal paradigm). In Brittonic, the meaning 'to stand' is expressed with the help of a denominative verb which eventually contains the same root *sta:-* (*steh₂*-), but underwent the part of speech change "verb → noun → verb". Because of this we mark Brittonic *sav-* with a separate index in the derivational drift-free dataset (Stage-2). The situation in Old Irish is different: in this language the meaning 'to stand' is rendered by a verbal noun *ses:ǝṽ* used in a construction 'he is in his standing'. Since this analytic pattern gradually replaces finite verb forms already from Old Irish onward and verbal noun is an integral part of Old Irish verbal paradigm, we see no reason in assigning a separate index to Old Irish *ses:ǝṽ* in the derivational drift-free dataset (Stage-2). As for the homoplasy-optimized dataset (Stage-3), there are two verbs in criss-crossed configuration: (1) *sta:-* (*steh₂*-) in the bulk of Inner IE branches, (2) *or-* (*h₃er-*) in Hittite and Albanian. The Hittite-Albanian match is a transparent case of homoplasy, since, first, the Proto-Inner IE meaning of *sta:-* (*steh₂*-) can be safely reconstructed as 'to stand', whereas the Proto-Inner IE meaning of *or-* (*h₃er-*) was likely 'to rise (move vertically)' as documented for its reflexes in Ancient Greek, Armenian, Latin, Old Indic. Second, the underlying meaning of Albanian *ri* ⟨*rri*⟩ is 'to be located' (hence the synchronous Albanian polysemy 'to sit / to stand'). In light of this we mark the Hittite and Al-

banian forms with two different indexes in the homoplasy-optimized dataset (Stage-3). NB: it is not excluded that at the Proto-IE (i.e., Indo-Hittite) level the verbs *sta:- (*steh₂-) and *or- (*h₃er-) formed the same aspectual opposition as *kʸey- 'to lie (stative)' vs. *legʰ- 'to lie down (change-of-state)' and *e:s- (*h₁eh₁s-) 'to sit, be sitting (stative)' vs. *sed- 'to sit down (change-of-state)'. I.e., the original meaning of *or- (*h₃er-) could be 'to stand up'.

80. 'star'.

Old Irish. *rʸ:e:d-glu* ⟨*rét-glu*⟩. A compound of rare *rʸ:e:d* ⟨*rét*⟩ 'star' with an uncertain second element. Alternative candidate: *rʸ:ind* ⟨*rind*⟩ 'constellation, star', may be applied to various kinds of celestial objects, including constellations.

Proto-Brittonic. **ster* is retained in all languages.

Proto-Germanic. **ster-n-o:n* is retained in all ancient and modern languages.

Proto-Slavic. **gvayzd-a:* ⟨**gvězda*⟩ is retained in all three subgroups.

Proto-East Baltic. **žvayžd-ya:* is retained in both Lithuanian and Latvian.

Proto-Iranian. **sta:r-* is retained in Avestan and all subgroups; this very stable term has direct Indo-Aryan and IE cognates.

Proto-Samoyed. **kɨnsV-kʋyǯ* (Helimski 1997: 278), retained in Mator, Kamass and Selkup, goes back to Proto-Uralic **kunʸsʸV* 'star'.

81. 'stone'.

Old Irish. *klox* ⟨*cloch*⟩. Secondary synonym: *lie* ⟨*lië*⟩ (disyllabic; later ⟨*lía*⟩) 'stone: (a) In general meaning; (b) Various specific meanings: whetstone; upper millstone; millstone; standing stone, pillar-stone, usually of a memorial stone, one marking place of death or burial; Ogham-stone'.

Proto-Brittonic. **maɣn* is retained as the main word for 'stone' in Cornish and Breton. The main Welsh word for 'stone' is ⟨*carreg*⟩, whose cognates in Cornish and Breton mean 'rock'. Welsh reflex of **maɣn* is ⟨*maen*⟩ 'stone, esp. one having some speciality or a particular use

(e.g. heavy stone used in athletic sports, grindstone, millstone)'. This meaning leaves little doubt that ⟨maen⟩ is the old generic word for 'stone', replaced by ⟨carreg⟩.

Proto-Germanic. *stayn-a-* is retained in all ancient and modern languages.

Proto-Slavic. *kaːm-uː*, obl. *kaːm-en-* ⟨*kamy, *kamen-*⟩ is retained in all three subgroups.

Proto-East Baltic. *ak-moː*, obl. *ak-men-* is retained in both Lithuanian and Latvian.

Proto-Iranian. *ac-man-* with a polysemy 'stone / heaven' is retained in Avestan, fully agreeing in form with the Old Indian word ⟨*áśman-*⟩ 'rock / stone'. The derived stem *ac-an-ga-* 'stone' < *ac-mn̥-ga-* is retained in all subgroups.

Proto-Samoyed. *pʋy* (Janhunen 1977: 112) is retained in Northern Samoyed, Kamass and Selkup.

82. 'sun'.

Tocharian B. We follow Lubotsky and Starostin (2003) and treat Proto-Tocharian *kaːwnə-* 'sun / day' as a borrowing from Proto-Turkic *gün* 'sun / day' (the source might be not proper Proto-Turkic, but its close extinct relative). Note that otherwise the Tocharian stem lacks any plausible etymology and that the polysemy 'sun / day' is atypical for the IE languages, being, on the contrary, characteristic of Turkic.

Old Irish. *gʸrʸiạn* ⟨*grían*⟩.

Proto-Brittonic. *hɔːwl* is retained in all languages.

Proto-Germanic. The Gothic paradigm (nom. ⟨*sauil*⟩, dat. ⟨*sunn-in*⟩) is likely to reflect the original Proto-Germanic heteroclitical declination: direct stem *soːw-el*, indirect stem *su-nn-*. However, even in Gothic this paradigm was already subject to analogical reformation (cf. the alternate nominative ⟨*sunn-o*⟩), and each of the other two branches generalized only one of the stems: the *-l*-stem becomes the norm in Scandinavian, the *-n*-stem - in West Germanic. Nevertheless, the root is consistently retained in all daughter languages.

Proto-Slavic. *suln-ik-a* ⟨*sъlnьko*⟩ is retained in all three subgroups.

Proto-East Baltic. *sawl-yaː* is retained in both Lithuanian and Latvian.

Proto-Iranian. The old heteroclitical paradigm *hwar / *hwan-, having direct Indo-Aryan and IE cognates, survives in this shape in Avestan. Later languages of all subgroups preserve forms derived from *hwar-.

Proto-Samoyed. *kʋyʋ (Janhunen 1977: 58) is retained as the word for 'sun' in all daughter languages save Selkup, where its meaning shifted to 'heat'. Related to Finno-Ugric verb *kaya- 'to appear, become visible'.

83. 'to swim'.

Old Irish. snaː-əðʸ ⟨snáïd⟩.

Proto-Brittonic. *nɔːṽ-iː-, retained in Welsh and Breton, is derived from a verbal noun *nɔːṽ. Cornish ⟨nija⟩ is related to Welsh ⟨neidio⟩ 'to jump; fly'.

Proto-Germanic. Not attested in Gothic; *swemm-an- is consistently suggested by data from the other two branches, although in modern Icelandic the word is generally replaced by ⟨synd-a⟩ of unclear origin.

Proto-Slavic. *ploː-, *pluʋ-, iterative plaːʋ-aː- ⟨*pluti, plavati⟩ is retained in all three subgroups.

Proto-East Baltic. *peld-eː is retained as 'to swim' in archaic Lithuanian and Latvian. The second candidate is *plawk-, reconstructed on the basis of the main Lithuanian verb 'to swim'. If the Yotvingian verb ⟨łaudt⟩ 'to swim' indeed directly corresponds to Lithuanian ⟨plaũkti⟩ 'to swim' (with the strange substitution of ⟨dt⟩ for *kt), it gives an advantage to plawk- over *peld-eː.

Proto-Iranian. *fraw- is retained in the meaning 'to swim' only in some Avestan texts, but direct Indo-Aryan and IE cognates prove its antiquity. In later Iranian languages, it was superseded by various other verbs (including *snaː- for which a Proto-Iranian meaning 'to bathe' can be reconstructed) or analytic expressions.

Proto-Samoyed. *u- (Janhunen 1977: 29), retained in Nenets, Mator and Selkup, goes back to Proto-Uralic *uyi- 'to swim'.

Etymological notes. Brittonic *nɔːṽ-iː- 'to swim' represents a denominative verb from the Celtic substantive *snaː-mu- 'swimming', which, in turn, goes back to the common Nuclear

IE verb *sna:- (*sneh₂-) 'to swim'. Because of the part of speech change, we mark the Proto-Brittonic form with a separate index in the derivational drift-free dataset (Stage-2).

84. 'tail'.

Ancient Attic Greek. *kérk-o-s* ⟨κέρκος⟩ 'tail' is a case when the Plato's lect possesses an innovative lexeme as compared to other Ancient Greek varieties (such as Homer or Herodotus) and Modern Greek dialects, where the basic term for 'tail' is ⟨οὐρή⟩. This, however, does not spoil the phylogeny since neither ⟨κέρκος⟩ nor ⟨οὐρή⟩ have cognate forms in the current dataset.

Old Irish. *er-bəl:* ⟨erball⟩. Apparently a compound with a second member *bal:* ⟨ball⟩ 'limb, member, organ (used of any portion of the body)'. The identity of the first member of the compound remains disputed.

Proto-Brittonic. *l:ost* is retained in Cornish and Breton, and (as a secondary synonym) in Welsh.

Proto-Germanic. Highly unstable. Not attested in Gothic; the most common Scandinavian equivalent is *hal-an-* (no parallels in other Germanic languages), the likeliest West Germanic candidate is *tagla-* (with the meaning 'tail' both in Old English and Old High German, but corresponding to Gothic ⟨tagl⟩ 'hair'). Since the generalization 'tail' > 'hair' is not highly likely, *hal-an-* remains as the only candidate securely projectable onto the Proto-Germanic level in the meaning 'tail'; however, in light of the total lack of stability of this word throughout the entire history of Germanic languages, it must be stated that, in all likelihood, we probably do not have any idea of what the generic 'tail' looked like in Proto-Germanic.

Proto-Slavic. An unstable item. Out of several mostly local terms, *xvast-u* ⟨*xvostъ⟩ is the best candidate, since it is attested in two subgroups: East (everywhere) and West (Old Czech, Old Polish, Slovak).

Proto-East Baltic. *o:deg-a:* is retained in Lithuanian and Latvian dialects.

Old Indic. We follow Lubotsky 2001: 312 and treat ⟨púcha-⟩ as a loan from an unknown source.

Proto-Iranian. **dum-a-* or suffixed **dum-b-a-* is retained in Avestan and all subgroups. Related to Old Indian *dumbaka-* 'a fat-tailed sheep'.

Proto-Samoyed. **tăɣwŏ* (Janhunen 1977: 150) is retained in all daughter languages save Enets.

85. 'that'.

Ancient Attic Greek. *ekêyno- ~ kêyno-* ⟨ἐκεῖνος ~ κεῖνος⟩ 'that' is most likely an inner Greek derivative from the locative adverb *ekêy ~ kêy* ⟨ἐκεῖ ~ κεῖ⟩ 'there, in that place' formed according to the typologically common pattern 'there' > 'that'. It is not excluded that initial *e-* or suffixal *-n-* are cognate with demonstrative roots in other Indo-European languages (e.g., with Slavic **an-* ⟨**on-*⟩ 'that', thus Sihler 1995: 390), but the main meaningful element in *ekêyno-* should be *-kêy-*.

Albanian. We follow some previous authors and treat Albanian *ay* ⟨ai⟩ 'that' as a continuant of IE **s-* with irregular loss of the sibilant in a grammatical word (the normal Albanian outcome of initial **s-* is ʒ- ⟨gj⟩ that coincides with the reflex of **y-*, medial **s > h > 0*).

Old Irish. The main opposition in Old Irish is between *so* ⟨so⟩ 'this' (phonetic variants ⟨sa, se, seo, sea⟩) and *sʸin* ⟨sin⟩ 'that'. These forms function as enclitics in the construction "article + noun + demonstrative enclitic": ⟨in fer so⟩ 'this man', ⟨in ben sin⟩ 'that woman'.

Proto-Brittonic. **hun:* (masculine), **hon:* (feminine) and **hɪn:* (neuter) are used in Welsh as a proximal pronoun 'this', while distal 'that' is derived from the same forms plus a second element of disputed origin: **hun:-e:ð, *hon:-e:ð, *hɪn:-e:ð*. The same system is reflected in Breton, so it can be safely reconstructed for Proto-Brittonic. Cornish and Breton also add suffixed locative adverbs ('here', 'there') to the same demonstrative stem.

Proto-Germanic. Precise onomasiological reconstruction is hindered here by the fact that the best reconstructible suppletive paradigm for Proto-Germanic deictic pronouns (masc. and fem. nom. **sa-*, neut. and indirect stem **θa-*) is semantically neutral in all ancient Germanic languages, while more specific proximal and distal forms are either formed from it with additional clitics or go back to other roots, but with weaker distribution in daughter languages. Without going into too much details, it seems best to resort to the following formal solution: (a) set up the suppletive variants **sa-* / **θa-* as corresponding to *both* the meanings 'this' and 'that'; (b) add Proto-Germanic **yayna-* 'that (yonder)' as a special addi-

tional pronoun for distal deixis, since it is clearly reconstructible based on data from all three branches.

Proto-Slavic. The Proto-Slavic system can be reconstructed as *s-i ⟨*sъ⟩ 'this' / *t- ⟨*tъ⟩ 'that (medial)' / *an- ⟨*on-⟩ 'that (distal)'. The proximal pronoun *s-i 'this' is attested in ancient languages: Old Church Slavonic, Old Russian, Old Novgorod, Old Czech, although in modern lects it is almost completely superseded with various affixal derivatives from *t- 'that'. The medial *t- 'that' is retained almost everywhere. The third member of the opposition, *an- ⟨*on-⟩ 'that (distal)', is attested in the South subgroup and some West lects (Sorbian), where the ternary system of the demonstratives is retained. In many Slavic lects, the original system got simplified into the binary one: 'this' / 'that'.

Proto-East Baltic. On the basis of the Lithuanian data, the Proto-East Baltic system can be reconstructed as *ši- 'this' / *ta- *that* (medial)' / *an-a- 'that (distal)'. In Latvian, it was simplified into the binary one with *an-a- omitted.

Proto-Iranian. On the basis of Avestan data, the Proto-Iranian system can be tentatively reconstructed as *ay- [masc.] ~ *i- [fem., neut.] 'this' / *ana- 'that' (medial) / *haw- [masc., fem.] ~ *awa- [neut.] 'that' (distal).

Proto-Samoyed. The reconstruction of Proto-Samoyed demonstratives is rather complicated. It involves at least three stems: *ta- ~ *tä- (Janhunen 1977: 150), *tə̆- (Janhunen 1977: 144) and *ti- (Janhunen 1977: 160–161). We tentatively reconstruct the basic opposition as that between distal *ta- ~ *tä- and proximal *tə̆-. Such an opposition is directly preserved in Kamass and Selkup, while Northern Samoyed languages suffered various restructurings of the system. It is possible, but not certain, that *ti- functioned as medial demonstrative. The Proto-Uralic deictic pronouns *te- 'this' and *to- 'that' (> Samoyed *tə̆-, *ta-) apparently have a common origin being cognate to Indo-European deictic *to-.

Etymological notes. It is hard to propose a reasonable reconstruction of deictic demonstratives for the Proto-IE or some intermediate levels. So we keep untouched potential cases of criss-crossing in the homoplasy-optimized dataset (Stage-3).

86. 'this'. See notes on 'that'.

87. 'thou'.

Old Irish. *tu:* ⟨*tú*⟩.

Proto-Brittonic. **ti:* is retained in all languages.

Proto-Germanic. The paradigm is well reconstructible as **θu:* [nom.] / **θe-* [obl.].

Proto-Slavic. The paradigm **tu:* ⟨**ty*⟩ [nom.] / **te-b-* ⟨**tebe*⟩ [obl.] is retained in all three subgroups.

Proto-East Baltic. The paradigm **tu:* [nom.] / **tav-* [obl.] is retained in both Lithuanian and Latvian.

Proto-Iranian. **tu:- ~ *tuw-am* is retained in Avestan and all subgroups. It has direct Indo-Aryan and IE cognates.

Proto-Samoyed. **tǎ-n* (Janhunen 1977: 147) is retained almost everywhere with the notable exception of (Tundra and Forest) Nenets, where 'thou' is etymologically 'thine body' and Forest Enets, where 'thou' is borrowed from Ket. Goes back to Proto-Uralic **ti-n* 'thou', which is a likely cognate of the Indo-European direct stem **ti* 'thou'.

88. 'tongue'.

Old Irish. *tʸeŋge* ⟨*tengae*⟩.

Proto-Brittonic. **tavɔːd* is retained in all languages.

Proto-Germanic. **tung-oːn* is retained in all ancient and modern languages.

Proto-Slavic. **enzuːk-u* ⟨**ęzykъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **inžuv-i-* is retained in Lithuanian, where this stem was phonetically accommodated to the verb **leyž-* 'to lick', in its original shape the stem is attested in Old Prussian.

Proto-Iranian. **hiʒ-u:-* (or **hiʒ-w-a:*) is retained in Avestan and all subgroups. It has direct Indo-Aryan and IE cognates.

Proto-Samoyed. **käǎ ~ *käǎɣ* (Janhunen 1977: 66), retained in all daughter languages except Tundra Nenets, goes back to Proto-Uralic **käli* 'tongue'.

89. 'tooth'.

Old Irish. *dʸeːd* ⟨*dét*⟩ 'tooth / set of teeth'. Secondary synonym: *fiạkəlʸ* ⟨*fíacail*⟩ 'tooth / tusk'.

Proto-Brittonic. **dant* is retained in all languages.

Proto-Germanic. **tunθ-u-* is retained in all ancient and modern languages.

Proto-Slavic. **zanb-u* ⟨**zǫbъ*⟩ is retained in all three subgroups.

Proto-East Baltic. **dant-i-* is retained in Lithuanian, its antiquity is proven by Old Prussian where it also means 'tooth' and external IE *comparanda*. The second candidate is **žamb-a-* which means 'sharp edge (of smth.), jag' in Lithuanian and 'tooth (anatomic); tooth, jag' in Latvian. The Latvian form *zuob-s* can regularly go back to Proto-Baltic **žamb-a-*, but its meaning 'tooth' suggests that actually it might be borrowed from or at least influenced by the East Slavic term for 'tooth' which has the shape like **zoːb-u* in the late 1st millennium BC (with **oː* regularly > Latvian *uo*). On the other hand, Finnic **hampas* 'tooth', borrowed from Baltic **žambas*, confirms the antiquity of the meaning 'tooth' in Baltic languages. The situation is somewhat similar to that of 'blood' (q.v.), where contact-driven developments are probable, although in the current case we are formally forced to treat **dant-i-* and **žamb-a-* as technical synonyms.

Proto-Iranian. The very stable stems **dant-* ~ **dant-an-* are retained in Avestan and almost all other subgroups. They have direct Indo-Aryan and IE cognates.

Proto-Samoyed. **timä* (Janhunen 1977: 163) is retained in all daughter languages.

Etymological notes. A difficult case with two stems in a criss-crossed configuration: **ədont-* ~ **dont-* (**h₁dont-*) and **gʸombʰ-o-*. Neither candidate has an advantage over another. In light of the **-yo-*derivatives meaning 'molar' in Greek and Old Indic, it is possible that **gʸombʰ-o-* originally meant 'molar', whereas **ədont-* ~ **dont-* denoted 'incisor', cf. the same lexical opposition without a single term for 'tooth' in Basque (for the alleged derivation from 'to eat' see the discussion in Kassian et al. 2015: 339). We are forced to keep this homoplasy untouched in the homoplasy-optimized dataset (Stage-3).

90. 'tree'.

Old Irish. *kran:* ⟨*crann*⟩.

Proto-Brittonic. **pren:* is retained in the meaning 'tree / wood' in Welsh, although the main synonym for 'tree' in Modern Welsh is ⟨*coeden*⟩ — a singulative from ⟨*coed*⟩ 'forest, wood, trees'. Cornish and Breton reflexes of **pren:* mean only 'wood', while the meaning 'tree' is expressed by the singulative of **gwıð* 'trees, forest'. We reconstruct **pren:* as the Proto-Brittonic word for 'tree' on the following grounds: 1) its Old Irish cognate ⟨*crann*⟩ means 'tree', and there is no other plausible candidate for the Proto-Celtic 'tree'; 2) reflexes of other Proto-Brittonic roots (**gwıð* 'trees, forest', **kɛːd* 'forest') mean 'tree' only in singulative form, whereas Welsh ⟨*pren*⟩ 'tree / wood' does not take a singulative suffix.

Proto-Germanic. **trew-a-* is retained as the basic word for 'tree' in Scandinavian languages; in Gothic, its reflex ⟨*triu*⟩ is only encountered in the meaning 'staff', whereas the basic meaning 'tree' is expressed by the innovation ⟨*bagms*⟩. In West Germanic, the situation is mixed: continental dialects mostly reflect **bagm-a-* (with the old root only preserved in derivates such as **terw-oːn* 'tar' > Dutch *teer*, etc.), whereas Old English has both ⟨*trēow*⟩ and ⟨*bēam*⟩, with unclear difference in usage. It may be speculated that Old English "revitalized" the original root under Scandinavian influence, despite **bagm-a-* 'beam, log' already functioning in Proto-West Germanic as the most neutral equivalent for 'tree (growing)'. Regardless of this speculation, however, Scandinavian evidence coupled with the external argument clearly indicates that **trew-a-* is eligible for Proto-Germanic as 'tree' although formally we may count both items as technical synonyms.

Proto-Slavic. **derv-a* ⟨**dervo*⟩ is retained in all three subgroups.

Proto-East Baltic. Not reconstructible. The Lithuanian word for 'tree' is derived from **medy-V* 'forest', which looks like an inner Lithuanian development. In Latvian, the meaning 'tree' for *kuok-s* was developed from 'a k. of log', which likewise looks like an inner Latvian development.

Old Indic. We follow Lubotsky 2001: 313 and treat ⟨*vr̥kṣá-*⟩ as a loan from an unknown source.

Proto-Iranian. **daːr-u-* 'tree / wood / stick' is retained in all subgroups. This word or its derivatives mean 'tree' in Southwestern and Northwestern Iranian languages. In light of external comparanda, **daːr-u-* can be safely posited as a Proto-Iranian term for 'tree'. Thus it

is likely that *da:r-u-* in the meaning 'tree' was superseded with *wan-a-* 'forest / (forest) tree' already in Avestan and eastern Iranian group.

Proto-Samoyed. *pa* (Janhunen 1977: 117), retained in all daughter languages, goes back to Proto-Uralic *pawi* 'tree'.

91. 'two'.

Old Irish. *da:* ⟨*dá*⟩.

Proto-Brittonic. *dɔ:w* is retained in all languages.

Proto-Slavic. *duv-a:* ⟨*dъwa*⟩ is retained in all three subgroups.

Proto-East Baltic. *dv-V-* is retained in both Lithuanian and Latvian.

Proto-Germanic. *two:u* is retained in all ancient and modern languages.

Proto-Iranian. *dwa-* is retained in Avestan and all subgroups. It has direct Indo-Aryan and IE cognates.

Proto-Samoyed. *kitä* (Janhunen 1977 71–72), retained in all daughter languages, goes back to Proto-Uralic numeral 'two'.

92. 'to go'.

Old Irish. *tʲe:dʲ* ⟨*téit*⟩. The present tense has two stems: *tʲe:dʲ* ⟨*téit*⟩ in 3 sg. and 2 pl. and *tʲiaɣ-* ⟨*tíag-*⟩ in the remaining forms. ⟨*tíag-*⟩ goes back to PIE root *steygʰ-*. The etymology of ⟨*téit*⟩ is disputed, but according to some versions, it can be derived from the same root *steygʰ-*. The verb also has suppletive future *rʲ:eɣ-* ⟨*reg-*⟩ and preterite *l:uðʲ* ⟨*luid*⟩ < PIE *lewdʰ-* (*h₁lewdʰ-*).

Proto-Brittonic. *aɣ-* is retained in all languages. The verb has suppletive subjunctive *el-* and verbal noun *mon-et*.

Proto-Germanic. *ge:-n-*, well attested in all three branches (in Gothic, only in its Crimean variety: *geen*), is clearly the best candidate for the neutral meaning 'to go' in Proto-Germanic; however, in many ancient and modern languages its usage visibly tilts to auxil-

iary mode (> 'to be going to', 'to intend to do smth.', etc.), leading to its replacement in the "physical" aspect by other stems (e.g. Gothic *ga=li:θ-an*; Old Norse *far-a* with its modern Scandinavian reflexes, etc.). The situation is further aggravated by the presence of two more suppletive stems with sufficient distribution to be projected onto the Proto-Germanic level: (a) **gaŋg-an-* (all three branches) and (b) **ei-* (the preterital forms *i-ddya* in Gothic and *e:o-de* in Old English). The semantic difference between this suppletive verb and **ge:-n-* is not easy to establish based on data from old texts, so all three may be included into comparison as synonyms on the Proto-Germanic level.

Proto-Slavic. **i:-d-* ⟨**id-*, **i-ti*⟩ is retained in all three subgroups. Distinct from **xad-i:-* ⟨**xodi-ti*⟩ 'to walk'.

Proto-East Baltic. **ey-* is retained in both Lithuanian and Latvian. Unlike the regular Lithuanian paradigm, the Latvian paradigm is suppletive with two different roots involved: **ey-* [pres.] / **ga:-* [pret.]. The Lithuanian continuant of **ga:-* is a regular verb meaning 'to rush, hurry'. No traces of **ga:-* can be found in Old Prussian and Slavic excepting some nominative stems. It is thus unclear from internal Baltic evidence whether this suppletion is to be projected onto the Proto-East Baltic level or it represents an inner Latvian introduction. External comparison (Indo-Iranian and Ancient Greek data), however, suggests that the Latvian suppletion could be an archaism, so we treat **ga:-* as a Proto-East Baltic preterite stem.

Proto-Iranian. **ay-* is retained in Avestan and all subgroups. It forms both present and aorist in Avestan. In many languages of all subgroups descendants of **čyaw-* 'to set in motion; to move' became the main verbs with the meaning 'to go'.

Proto-Samoyed. **men-* (Janhunen 1977: 94), retained in Nenets, Nganasan and Kamass, goes back to Proto-Uralic **meni-* 'to go'.

Etymological notes. There is an etymological match between Ancient Greek *elt^h-* ⟨ἐλθ-⟩ and Old Irish *l:uð^y* ⟨*luid*⟩; both roots function as a suppletive aorist. This can be a chance coincidence, but on the other hand, it is not excluded that the Greek-Irish match reflects an old suppletive paradigm. Because of this, just as in the similar case with 'to come' (q.v.), we keep the same index for *elt^h-* and *l:uð^y* in the homoplasy-optimized dataset (Stage-3).

93. 'warm'.

Old Irish. *tʸee* ⟨teë⟩. Goes back to **teent-* < **tepent-* from PIE root **tep-* 'to be warm'.

Proto-Brittonic. **teːmː*, retained in Breton, Cornish and Southern Welsh, goes back to **teemmo-* < **tepesmo-* from PIE root ⟨**tep-*⟩ 'to be warm'.

Proto-Germanic. The opposition **war-ma-* 'warm' : **xayt-a-* 'hot' is well reconstructible for Proto-Germanic and is preserved in many West Germanic and Scandinavian languages (in Gothic, these adjectives are not attested), although in some modern Scandinavian languages there are signs of merger of both meanings within reflexes of **war-ma-*.

Proto-Slavic. **tep-l-* ⟨**tepl-*⟩ is retained in all three subgroups, meaning 'warm'. Distinct from **gar-entj-* ⟨**gorętj-*⟩ 'hot', active participle from **gar-eː-* ⟨**gorěti*⟩ 'to burn (intr.)'.

Proto-East Baltic. **šil-t-a-* 'warm' is retained in both Lithuanian and Latvian. Participle from **šil-* 'to become warm'. Distinct from **karš-t-a-* 'hot', participle from **karš-* 'to become hot'.

Proto-Iranian. **gar-ma-* 'hot / warm' is retained in Avestan and all subgroups; it has direct Indo-Aryan and IE cognates.

Proto-Samoyed. There are two candidates. The first, **yu-pʋ* (Janhunen 1977: 47–48), is derived from the verb **yu-* 'to be warm / to melt'. This is the main word for 'warm' in Nenets, Enets and Mator. The second candidate, **päywä* (Janhunen 1977: 120), means 'heat, warmth' in Nganasan and 'warm' in Selkup. Its Finnic and Saami cognates (< Proto-Uralic **päywä*) mean 'sun, day'.

Etymological notes. The bulk of the forms involved reflects the adjectival derivative from the verb **gʰʷer-* 'to be warm' with the suffix **-mo-*: Ancient Greek *tʰer-mó-*, Armenian *ǰɛɹ-m*, Germanic **war-ma-*, Iranian **gar-ma-*. A different morphological pattern is observed in Albanian, where *n=grɔh-tə* is a synchronic participle from an Albanian verb (which indeed goes back to **gʰʷer-*); because of this we mark the Albanian form with a separate index in the derivational drift-free dataset (Stage-2). A similar case is Old Irish *tʸee*, Brittonic **teːmː*, Slavic **tep-l-* which represent different adjectival formations from the verbal root **tep-* 'to be warm (*vel sim.*)': **-ent-* in Old Irish, **-esmo-* in Brittonic and **-lo-* in Slavic. Because of this we assign three different indexes to the Old Irish, Brittonic and Slavic forms in the derivational drift-free dataset (Stage-2).

94. 'water'.

Old Irish. *us^ʸk^ʸe* ⟨*uisce*⟩ goes back to **ud-sk-yo-* from the same root as PIE **wod-ṛ* 'water'. There are indications, however, that this word is not a direct continuation of the PIE word for 'water'. First, the Old Irish word has no traces of the heteroclitic suffix **-r/-n-*, being formed directly from the root **wed-*. Second, Old Irish has another word for 'water', ⟨*dobur*⟩, related to Proto-Brittonic **duvr*. This word is found only in glossaries, compounds like ⟨*doborchú*⟩ 'otter' (lit. 'water-dog') and place names. It is possible that ⟨*uisce*⟩ is an innovation that replaced ⟨*dobur*⟩. The latter word is itself a Common Celtic replacement of PIE **wod-ṛ*.

Proto-Brittonic. **duvr* is retained in all languages.

Proto-Germanic. The heteroclitical paradigm **wat-ar*, indirect stem **wat-an-* is reconstructible for Proto-Germanic, although West Germanic languages have generalized the former stem, and Scandinavian languages and Gothic the latter.

Proto-Slavic. **vad-a:* ⟨**voda*⟩ is retained in all three subgroups.

Proto-East Baltic. **vand-o:*, obl. **und-en-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **a:p-* / **ap-* is retained in Avestan and almost all other subgroups. A very stable term; its Old Indian cognate is used mostly in plural and means in that case 'the Waters considered as divinities'.

Proto-Samoyed. **wet* (Janhunen 1977: 176), retained in all Samoyed languages, goes back to Proto-Uralic **weti* 'water', which is a likely cognate of Indo-European **wed-* 'water' (Illich-Svitych 1967: 334).

Etymological notes. Old Irish *us^ʸk^ʸe* (presumably < **ud-sk-yo-*) may indeed contain the common IE root **wed-/*wod-* 'water', but the complex suffixal pattern suggests that at best we are dealing here with a deverbative stem; because of this we mark the Old Irish form with a separate index in the derivational drift-free dataset (Stage-2).

95. 'we'.

Old Irish. *s^ʸn^ʸi:* ⟨*sní*⟩.

Proto-Brittonic. **ni:* is retained in all languages.

Proto-Germanic. The suppletive paradigm *wiːz [nom.] / *uns [obl.] is retained in all ancient and modern languages. Of special note is the separate dual stem *wit 'we two', well attested in most ancient languages but largely out of usage in modern ones (although cf. Icelandic ⟨við⟩, Faroese ⟨vit⟩ 'we', continuing the old dual rather than plural form); however, it clearly has the same root as *wiːz.

Proto-Slavic. The suppletive paradigm *muː ⟨*my⟩ [nom.] / *naːs- ⟨*nasъ⟩ [obl.] is retained in all three subgroups.

Proto-East Baltic. The paradigm *mes ~ *meːs [nom.] / *mus- ~ *muːs- [obl.] is retained in both Lithuanian and Latvian.

Proto-Iranian. The suppletive paradigm *way-am [nom.] / *ah-ma- [obl.] is retained in this shape in Avestan. In later languages, the paradigm is levelled in favor of oblique cases, usually the genitive with the metathesis *Vmaːx- < *amaːxam < *ahmaːkam.

Proto-Samoyed. *me- (Janhunen 1977: 91), retained in Nganasan, Mator, Kamass and Selkup, goes back to Proto-Uralic *me- 'we'. Nenets and Enets replaced the original pronoun by dual/plural forms of 'I'.

Etymological notes. The Armenian and Balto-Slavic forms with *m- for expected *w- most likely have been independently carried over by analogy either from the verbal 1st p. pl. ending or the 1st p. sg. pronoun or both.

96. 'what'.

Old Irish. kʲið ~ kʲeð ⟨cid ~ ced⟩.

Proto-Brittonic. *pɪ ~ *pa. The meaning 'what?' is usually expressed in Brittonic languages by a collocation 'which thing?', where *pɪ ~ *pa is an adjectival interrogative pronoun 'which'. Different languages use different words for 'thing' in this construction, so that only *pɪ ~ *pa can be reconstructed for Proto-Brittonic.

Proto-Germanic. *xwa-t (with the old neuter ending) is retained in all ancient and modern languages.

Proto-Slavic. *č- ⟨*čьto⟩ is retained in all three subgroups.

Proto-East Baltic. *ka- is retained in both Lithuanian and Latvian.

Proto-Iranian. The Proto-Iranian paradigm of the interrogative pronoun 'who, what' is likely to be reconstructed as m. *ka-s*, f. *kaː*, n. *či-t* (cf. the same distribution in the Proto-Slavic paradigm m.-f. *ku-ta* ⟨*kъto⟩, n. *či-ta* ⟨*čьto⟩). It is at least partially retained in Avestan and Old Persian, although already in Avestan n. *či-t* tends to be superseded with probably secondary *ka-t*, whereas m. *ka-s* can be superseded with probably secondary *či-s* and so on.

Proto-Samoyed. *mə* (Janhunen 1977: 91), retained as an interrogative pronoun 'what' in Enets and Nganasan, goes back to Proto-Uralic *mV* 'what?'.

97. 'white'.

Old Irish. *fʲind* ⟨find⟩. Secondary synonyms: *baːn* ⟨bán⟩ 'white, fair, bright; pale', *gʲel* ⟨gel⟩ 'fair, white, bright, shining'.

Proto-Brittonic. *gwɪnː* is retained in all languages.

Proto-Germanic. *xwiːt-a-* is retained in all ancient and modern languages.

Proto-Slavic. *beːl-* ⟨*běl-⟩ is retained in all three subgroups.

Proto-East Baltic. *bal-t-a-*, retained in both Lithuanian and Latvian, is a participle of *bal-* 'to become white'.

Proto-Iranian. *cwayt-a- ~ *cwit-a-* is retained in Avestan and all subgroups; this stem has Indo-Aryan cognates with meanings 'white / bright / light'.

Proto-Samoyed. *sər* (Janhunen 1977: 138) is retained in all daughter languages save Mator. Possibly connected to Proto-Samoyed *sər* 'ice'.

Etymological notes. Slavic *beːl-* and East Baltic *bal-t-a-* are probably related, although have different derivational history: the Slavic form is a primary adjective, the Baltic one is a participle of the Baltic verb *bal-* 'to become white'. Because of this we assign different indexes to the Slavic and Baltic forms in the derivational drift-free dataset (Stage-2).

98. 'who'.

Old Irish. *kʲia̯* ⟨cía⟩.

Proto-Brittonic. **pe:* is retained in all languages.

Proto-Germanic. **xwa-z ~ *xwi-z* (masculine), **xw-o:* (feminine) is well reconstructible for the proto-level; reflexes of the root are found in all ancient and modern languages.

Proto-Slavic. **k-* ⟨**kъto*⟩ is retained in all three subgroups.

Proto-East Baltic. **ka-* is retained in both Lithuanian and Latvian.

Proto-Iranian. **ka-* is retained as interrogative 'who?' in Avestan, Old Persian and all other subgroups (frequently in the form of the genitive singular **ka-hya-*). Further see notes on 'what?'.

Proto-Samoyed. **ke-* (Janhunen 1977: 69), retained in Northern Samoyed, Mator and Kamass, goes back to Proto-Uralic **kV-* 'who?', which is a likely cognate of Indo-European **kʷi-* 'who?'.

99. 'woman'.

Old Irish. *bʸen* ⟨*ben*⟩.

Proto-Brittonic. **ben* is retained as the main word for 'woman' in Cornish. Breton preserves this word as a secondary synonym. South Welsh has ⟨*menyw*⟩ — an analogically altered form of ⟨*benyw*⟩, derived from the same root.

Proto-Germanic. **kwen-o:n* is well attested in most ancient languages and preserved in modern Scandinavian ones; elsewhere, largely replaced by innovations (Old English ⟨*wíf*⟩, German ⟨*Frau*⟩, etc.) and either lost or semantically shifted (e.g. to English ⟨*queen*⟩).

Proto-Slavic. **žen-a:* ⟨**žena*⟩ or its suffixed derivatives are retained in all three subgroups.

Proto-East Baltic. Not reconstructible. The Lithuanian word for 'woman' is a derivative from 'mother'. In Latvian, 'woman' is derived from 'wife'. The original Proto-Baltic term **gen-a:* 'woman' is only retained in Old Prussian.

Proto-Iranian. **ǯan-i-* is widely attested as a basic term for 'woman' in Iranian languages (e.g., in Middle Persian) and is a direct continuant to PIE **gʷen-*, which is indubitably the main equivalent for 'woman' on the PIE level. In Avestan, this is not the basic term, but nevertheless it is still attested in Avestan in the collocation 'woman (and) man' as well as

with the meaning 'wife' (semantic shift 'woman' > 'wife' is very common cross-linguistically). The Avestan basic terms are probably *naːr-iː-* ⟨*nāirī-*⟩ and *naːr-i-k-aː-* ⟨*nāirikā-*⟩ 'woman', a feminitive from 'man', but this looks like an Avestan innovation (thus probably either morphological or semantic parallel introductions in Avestan and Vedic, cf. Vedic *nˈaːr-i-* ~ *nˈaːr-iː-* 'woman'). Proto-Iranian *str-iː-* is likely to be reconstructed with the meaning 'female (subst.)', although, e.g., in Khotanese Saka it became a main term for 'woman'.

Proto-Samoyed. *ne* (Janhunen 1977: 100), retained in all Samoyed languages, goes back to Proto-Uralic *näɣi* 'woman'.

100. 'yellow'.

Old Irish. *buðʸe* ⟨*buide*⟩.

Proto-Brittonic. *mel-ɪn*, derived from *mel* 'honey', is retained in all languages.

Proto-Germanic. *gelw-a-* ~ *gulu-* is well attested in all branches, except for Gothic, and is surprisingly stable in modern languages.

Proto-Slavic. *žilt-* ⟨*žьlt-*⟩ is retained in all three subgroups.

Proto-East Baltic. *gelt-a-* is retained in both Lithuanian and Latvian (in modern languages, various suffixed derivatives from this root are more commonly used for 'yellow').

Proto-Iranian. *ʒar-i-* with polysemy 'green / yellow' or the suffixed derivative *ʒar-i-ta-* 'green' are retained in this meaning in Avestan and all subgroups; these stems have Indo-Aryan and IE cognates.

Proto-Samoyed. *tʋsV-* ~ *çʋsV*. The word for 'yellow' can only be reconstructed for Northern Samoyed.

Etymological notes. Several cognate forms involved represent adjectives with the root *g^{yh}elə-* (*g^{yh}elh₃-*) modified with adjectival suffixes: *-u-* (Germanic), *-i-* (Iranian) or *-to-* (Balto-Slavic, Indo-Iranian). But since there is no evidence for the verbal nature of *g^{yh}elə-* (*g^{yh}elh₃-*), we mark these adjectives with the same index in the derivational drift-free dataset (Stage-2).

101. 'far'.

Old Irish. *kʸiạn* ⟨*cían*⟩.

Proto-Brittonic. **pel:* is retained in all languages.

Proto-Germanic. The adverbial form **ferr-ay* is retained in all ancient and modern languages.

Proto-Slavic. **daːl-ek-a* ⟨**daleko*⟩ is retained in all three subgroups. Derived from **daːl-i* ⟨**dalь*⟩ 'long distance, expanse'.

Proto-East Baltic. Adverbial **taːl-V* is retained in both Lithuanian and Latvian.

Proto-Iranian. **duːr-ay* is retained in Avestan and all subgroups; it has Indo-Aryan and IE cognates.

Proto-Samoyed. **kuntə̆-kʋ* (Janhunen 1977: 78), retained in all languages save Nenets, is derived from **kuntə̆* 'long, length'.

Etymological notes. There are two roots in criss-crossed configuration: (1) **dwaː-* (**dweh₂-*) in Hittite and Indo-Iranian; (2) **per-* in Armenian and Germanic. The first one seems more archaic in the meaning 'far', nevertheless in the absence of reliable evidence we prefer to keep this case of homoplasy untouched in the homoplasy-optimized dataset (Stage-3).

102. 'heavy'.

Old Irish. *trom* ⟨*tromm*⟩.

Proto-Brittonic. **trum* is retained in Welsh. Cornish and Breton have innovations based on different words for 'weight' eventually borrowed from Latin.

Proto-Germanic. The situation here is complex, since this meaning is quite unstable in Germanic languages. On one hand, Gothic ⟨*kaurus*⟩ 'heavy', without any parallels in the other two branches, looks like an excellent candidate for Proto-Germanic status because of the external argument (obvious cognacy with the corresponding words meaning 'heavy' in Sanskrit, Greek, etc.); it must, however, be noted that the adjective itself is attested only once in the Gothic corpus and in a figurative meaning at that ('weighty', said of letters). On the other hand, there is also Proto-Germanic **sweːr-a-*, clearly the main equivalent for

'heavy' in Proto-West Germanic (Old English, Old High German, etc.) and also attested in both of the other branches — as Old Norse ⟨svár-r⟩ 'heavy, grave' (an obsolete archaic poetic word; the neutral equivalent for 'heavy' in Proto-Scandinavian was already the innovation *θung-a-) and as Gothic ⟨swer-s⟩ 'esteemed, dignified' (an easy potential semantic shift from 'heavy'). Without the external argument, *swe:r-a- would have been the logical choice for inclusion, but in the light of its existence it is probably more correct to include both roots on the list as technical synonyms.

Proto-Slavic. *tenž-ik- ⟨*tęžьk-⟩ is retained in all three subgroups. Derived from *teng-a: ⟨*tęga⟩ 'traction; heavy weight'.

Proto-East Baltic. *smag-u- is retained as 'heavy / difficult' in both Lithuanian dialects and Latvian.

Proto-Iranian. *gar-u- is retained in all subgroups; this stem has direct Indo-Aryan and IE cognates.

Proto-Samoyed. *säç- (Janhunen 1977: 139) is retained in Northern Samoyed, Mator, Kamass and Selkup.

Etymological notes. Tocharian B kramar-cc-e eventually contains the same root as the common Inner IE adjective $*g^w\dot{r}$-u- ($*g^w\dot{r}h_2$-u-) 'heavy', although it is necessary to postulate a complex derivational chain with different parts of speech involved: 'heavy' → 'weight, heaviness' → 'heavy'. Because of this we mark the Tocharian form with a separate index in the derivational drift-free dataset (Stage-2).

103. 'near'.

Old Irish. $ag^w\partial s$ ~ $og^w\partial s$ ⟨acus ~ ocus⟩.

Proto-Brittonic. *agos ~ *ogos is retained in Welsh and Cornish. This word is traditionally considered a borrowing from Old Irish ⟨acus ~ ocus⟩ due to irregular correspondence of internal consonants and the fact that Welsh also has an adjective ⟨wng⟩ 'near, close; nearness, vicinity' that can go back to Proto-Celtic *onko- (> Old Irish ⟨acus ~ ocus⟩).

Proto-Germanic. The adverbial form *ne:xw-a, or one or more of its derivates, are retained in all ancient and the majority of modern languages.

Proto-Slavic. *bliːz-a ⟨*blizo⟩ or diminutive *bliːz-uk-a ⟨*blizъko⟩ are retained in all three subgroups.

Proto-East Baltic. *tuv-i, which is retained as temporal 'right now' in Lithuanian and as spatial 'near' in Latvian, is the best candidate.

Proto-Iranian. Arverbs 'near', derived from *nazd-, are attested in all subgroups. Note Avestan *asn-ay* < *n̥zd-na-.

Proto-Samoyed. *wan-i- (Helimski 1997: 301) is retained in Enets, Mator and Kamass.

104. 'salt'.

Old Irish. *salən:* ⟨salann⟩.

Proto-Brittonic. *haleːn is retained in all languages.

Proto-Germanic. *salt-a- is retained in all ancient and modern languages.

Proto-Slavic. *sal-i ⟨*solъ⟩ is retained in all three subgroups.

Proto-East Baltic. *saːl-i- is retained as 'salt' in Latvian and in a derivative stem for 'pickle' in Lithuanian.

Proto-Iranian. *namada-ka- ~ namad-ka- is attested in all subgroups (probably including Avestan). This word can be analysed as *na=mad-a-ka- 'not (having) moist' or, otherwise, it could be derived from *nab- / *nam- 'to be wet'. Kurdish ⟨xwē⟩ 'salt' and Balochi ⟨vād⟩ are derived from *hwaːd- 'delicious; to be delicious'.

Proto-Samoyed. *sər (Janhunen 1977: 138), the Proto-Samoyed word for 'ice', which in Northern Samoyed languages also means 'salt'. Mator and Kamass words for 'salt' are borrowed from Turkic, Selkup word for 'salt' is apparently an Iranian loan.

105. 'short'.

Old Irish. *gʸer:* ⟨gerr⟩.

Proto-Brittonic. *bɪr: is retained in all languages.

Proto-Germanic. Not properly reconstructible; there is not a single root that would function as the basic equivalent for 'short' in at least two of the subgroups, and the external argument is largely inapplicable to all the candidates. For Proto-West Germanic, the best option is *skurt-a-* (related to *skert-an-* 'to cut off, trim, shorten'); for Proto-Scandinavian, it is probably *stutt-a-* < *stunt-a-* < *stun-yan-* 'to trim, shorten', without further etymology.

Proto-Slavic. *kart-uk-* ⟨*kortъk-⟩ is retained in all three subgroups.

Proto-East Baltic. *iːnš-a-* is retained both in Lithuanian dialects and in Latvian.

Proto-Iranian. *mr̥ʒ-u-* is retained in Avestan (in temporal sense), Khotanese and Sogdian. External comparanda prove its antiquity.

Proto-Samoyed. *kăym* (Janhunen 1977: 51) is retained in all Samoyed languages.

106. 'snake'.

Old Irish. *naθʸərʸ* ⟨nathir⟩.

Proto-Brittonic. *nadr* is retained in all languages.

Proto-Germanic. Reconstruction is difficult, because of polysemous relations with the meaning 'worm' and multiple euphemistic replacements in various subgroups. Still, the stem *wurm-i-* is probably the best candidate, since it is explicitly attested in the meaning 'snake' in all the ancient languages of all three subgroups (in Gothic exclusively so; in Old Norse, Old English, Old High German etc. this seems to be a generic term to describe all crawling species — snakes, worms, even dragons). Other candidates are all flawed one way or another: thus, *nadr-a-* more frequently refers to specific subtypes of snakes ('viper', 'adder'), *slang-an-* is clearly derived from the verb *sling-an-* 'to creep, to slink', *sneːk-an- ~ *snak-an-* is derived from *snak-an-* 'to crouch' etc. Of note, perhaps, is Old High German ⟨unk⟩ 'snake, lizard', allegedly continuing PIE *h₂engʷʰ-*, but the form, only attested in one medieval form of one West Germanic language, raises significant doubts about its origins (not to mention that, according to Kroonen 2013: 560, it does not even properly correspond to its correlates in Baltic and Armenian).

Proto-Slavic. A complicated case. The masculine noun *zm-iy-i* ⟨*zmьjь⟩, derived from Proto-Slavic *zem-i* ⟨*zemь⟩ 'earth' (q.v.), can be safely reconstructed with the meaning '(large mythological) serpent' on the basis of evidence from all three groups. Its feminine deriva-

tive *zm-iy-a: ⟨*zmьja⟩ means 'snake (in general)' in the East and South subgroups, and specifically 'viper' in the West subgroup. The formal distribution points to *zm-iy-a: as a Proto-Slavic term for 'snake'. On the other hand, there exists the non-derivative noun *anž-i ⟨*ǫžь⟩ which means 'snake' in the West subgroup and 'grass-snake, innocuous snake' in the East and South subgroups ('slow-worm, legless lizard' in Serbian). The distribution of *anž-i is thus weaker than that of *zm-iy-a:, but *anž-i has the direct Baltic cognate *ang-i- 'snake'. We prefer to treat *zm-iy-a: and *anž-i as technical synonyms for Proto-Slavic.

Proto-East Baltic. *ang-i- is retained as 'snake / viper' in Lithuanian (not the main word for 'snake' in modern Lithuanian) and as 'viper' in Latvian. The antiquity of *ang-i- is proven by its meaning 'snake' in Old Prussian. The basic Lithuanian term for 'snake' is based on the verbal stem *gi:v-a- 'to live'. In Latvian, superseded with the obscure word čusk-a 'snake'.

Proto-Iranian. *až-i- is retained in Avestan, in Middle Persian in the meaning 'dragon', in Khwarezmian and Munji; this stem has Indo-Aryan and IE cognates, which proves that it should be considered the main Iranian word for 'snake'. A stem *ma:r-a- probably derived from *mar- 'to kill' serves as a designation of 'snake' in Southwestern and Northwestern Iranian languages. In many Iranian languages, especially Eastern, descendants of *kr̥m-i- 'worm' altered the meaning to 'snake' (or the polysemy 'worm / snake' can be found). Eastern languages also have one stem without a secure reconstruction, whose reflexes are Shughni ⟨divūsk⟩, Sarikoli ⟨tъɪfъɪsk⟩, Wakhi ⟨fuks⟩ 'snake' etc.

Proto-Samoyed. *ki-wä (Janhunen 1977: 72), retained as a word for 'snake' only in Selkup, goes back to Proto-Uralic *küyi-wä 'snake' (Aikio 2002: 43–44).

Etymological notes. There are two phonetically similar stems in criss-crossed configuration. (1) *og^{wh}i- (*h₃eg^{wh}i-) in Greek and Indo-Iranian. (2) *ang^{wh}i- (*h₂eng^{wh}i-) in Armenian, Latin and Balto-Slavic. It is hard to make a reasoned decision which root represented a basic term for 'snake' at the Proto-Inner IE level. We are forced to keep this case of homoplasy untouched in the homoplasy-optimized dataset (Stage-3).

107. 'thin'.

Old Irish. tane ⟨tanae⟩ 'thin / slender / narrow'. Secondary synonym: koi̯l ⟨cóil⟩ 'thin / slender / narrow'. Unlike ⟨tanae⟩, ⟨cóil⟩ can be applied to human body.

Proto-Brittonic. **tanaw ~ *tenew* is retained in all languages.

Proto-Germanic. **θunn-u-* is retained in all ancient and modern languages.

Proto-Slavic. **tin-uk-* ⟨**tьnъk-*⟩ is retained in all three subgroups, meaning 'thin 2D/1D'.

Proto-East Baltic. **tenv-a-* is retained as 'slender, svelte' in Lithuanian and as generic 'thin 2D/1D' in Latvian. In Lithuanian, it was superseded with the new opposition 'thin 2D' / 'thin 1D' based on the stems **plaː-n-a-* 'flat' and **layb-a-* 'lean, weak' respectively. An alternative solution is to reconstruct the original Proto-East Baltic system as a binary one: **tenv-a-* 'thin 2D' / **layb-a-* 'thin 1D', but it seems a less parsimonious scenario in light of the Proto-Slavic cognate adjective **liːb-* 'weak'.

Proto-Iranian. **tan-u-ka-* is retained in the majority of subgroups; this stem has direct Indo-Aryan and IE cognates. The stem **naːz-u-ka-*, originally meaning 'tender', replaced it in some Southwestern and Northwestern languages.

Proto-Samoyed. **yʋptʋ* (Janhunen 1977: 38) is retained in Northern Samoyed, Mator and Selkup.

108. 'wind'.

Old Irish. *gaįθ* ⟨*gaíth*⟩.

Proto-Brittonic. **gwɪnt* is retained as the main word for 'wind' in Welsh and Cornish. Replaced in Breton by ⟨*avel*⟩ 'wind', whose Welsh cognate ⟨*awel*⟩ means '(light) wind, breeze'. Etymologically, these words go back to the same PIE root **xweː-* (**h₂weh₁-*): **gwɪnt* < **h₂weh₁-nt-o-*, **awel* < **h₂ewh₁-el-*.

Proto-Germanic. **wend-a-* is retained in all ancient and modern languages.

Proto-Slavic. **veː-tr-u* ⟨**větrъ*⟩ is retained in all three subgroups. Derived from the verb **veː-yaː-* ⟨**vějati*⟩ 'to blow'.

Proto-East Baltic. **weː-y-a-* is retained in both Lithuanian dialects and Latvian.

Proto-Iranian. **waː-ta-* is retained in Avestan and all subgroups; this stem has direct Indo-Aryan and IE cognates.

Proto-Samoyed. **märkä* (Janhunen 1977: 93) is retained in all daughter languages save Nganasan.

Etymological notes. The bulk of the involved forms reflects the participle in **-nt-* from the verb **xwe:-* (**h₂weh₁-*) 'to blow (of wind)'. Slavic **ve:-tr-u* and Baltic **we:-y-a-* represent deverbative stems from the same root, but with other suffixal patterns. Because of this we assign separate indexes to the Slavic and Baltic forms in the derivational drift-free dataset (Stage-2).

109. 'worm'.

Old Irish. *kruṽʸ* ⟨*cruim*⟩.

Proto-Brittonic. **prɪṽ* is retained in all languages as a generic term for worms and similar creatures, cf. the meaning of Welsh ⟨*pryf*⟩ 'insect, esp. (house)fly, also of other similar animals; larva of various insects, maggot, grub, (earth)worm, also of other similar animals'. The names for 'earthworm' in Northern Welsh and Breton are compounds or collocations with **prɪṽ*.

Proto-Germanic. See 'snake' above on notes about **wurm-i-* as the original generic term covering all crawling species. However, a more specific root with the semantics of 'worm / maggot / moth' is **maθ-o:n*, well attested in most ancient and modern languages (often in derived variants, e.g. **maθ-ik-an-* > Old Norse ⟨*maðk-r*⟩). It is not entirely certain if this term could specifically denote 'earthworm' in Proto-Germanic, but it does feature in compound formations with this meaning in modern Scandinavian languages at least. Finnic **mato* 'worm' is possibly borrowed from Germanic source (an alternative version is that the Finnic word is cognate to Proto-Saami **muoce:* 'moth').

Proto-Slavic. **čirv-i* ~ **čirm-i* ⟨**čьrvъ, *čьrmъ*⟩. The stem **čirv-i* as a generic designation of 'worm' is attested almost everywhere including ancient languages, the same concerns various derivatives with the root **čirv-* (e.g., **čirv-i:-* 'to make red'). On the contrary, the root **čirm-* is only known from several derivatives whose meanings are connected with 'red', not 'worm' *per se* (including **čirm-in-* ⟨**čьrmьn-*⟩ 'red' q.v.). The external comparison strongly suggests that **čirm-* must be an original shape of the root for 'worm'. Thus the most likely scenario is that **čirm-i* 'worm' occasionally dissimilated > **čirv-i* already in Proto-Slavic (e.g., due to influence on the part of **morv-i* 'ant'), having survived with etymo-

logical -m- in a few derived stem with the secondary color semantics. Also see notes on 'red'.

Proto-East Baltic. *sleyk-a- is attested as 'earthworm' in both Lithuanian and Latvian. Its antiquity is proven by the Old Prussian word for 'earthworm' which goes back to *slayk-a- (with another ablaut grade). Apparently *sleyk-a- ~ *slayk-a- originates from the meaning 'slimy'. The old term *kirm-i- is retained as generic 'worm' in Lithuanian (but not for 'earthworm') and in many derivative with the worm/maggot semantics in Latvian.

Proto-Iranian. *kr̥m-i- is retained in Avestan and the majority of other subgroups; this stem has Indo-Aryan and IE cognates.

Proto-Samoyed. *çuk ~ *çukə (Janhunen 1977: 34) is attested as the main word for 'worm' and 'insect' in Mator and Selkup. Its Northern Samoyed cognates mean rather 'fly / larva of a fly'. The main word for 'worm' in Northern Samoyed, *kə̆lə̆-, lacks cognates in Southern languages. We list both words as synonyms.

110. 'year'.

Old Irish. bli̯aδ-ənʸ ⟨blíadain⟩.

Proto-Brittonic. *ble:δ or its derivatives are retained in all languages.

Proto-Germanic. *ye:r-a- is retained in all ancient and modern languages.

Proto-Slavic. There are two main candidates intertwined within the subgroups or even co-existing in a single language: *gad-u ⟨*godъ⟩ and *le:t-a ⟨*lěto⟩. We are forced to treat them as synonyms.

Proto-East Baltic. *met-a- is retained as 'year' in Lithuanian and as 'time span' in Latvian. Its antiquity is proven by the meaning 'year' in Old Prussian.

Proto-Iranian. According to Avestan data and IE cognates, *ya:r- meant 'year' in Proto-Iranian. *car-da- with the original Proto-Indo-Iranian meaning 'autumn' alters it to 'year (of somebody's age)' in Avestan and to 'year (in general)' in later Iranian languages.

Proto-Samoyed. *pŏ (Janhunen 1977 127) is retained in all daughter languages.

Etymological notes. There are two roots in criss-crossed configuration. (1) **wet-* or its derivative **wet-os-* in Hittite, Ancient Greek (outside our dataset), Albanian, Old Indic. (2) Heteroclite **yeː-r ~ *yoː-r*, obl. **yeː-n-* (**yeh₁-r*) and its derivatives which mean 'year' in Germanic and Iranian and 'time', 'season', 'hour', 'spring', 'lamb' in Anatolian, Ancient Greek, Balto-Slavic. It seems probably that the original meaning of **yeː-r* was something like 'time period' or 'season', i.e., that the meaning 'year' in Germanic and Iranian represents parallel semantic development. On the other hand, it is not excluded that both terms coexisted with the meaning 'year' at the Proto-IE level, e.g., the first one denoted 'year (as a time period)', the other one denoted 'year (as an age marker)'. We prefer keep this case of homoplasy untouched in the homoplasy-optimized dataset (Stage-3).

Overview of lexical innovations in some Inner IE clades (Greek-Armenian, Balto-Slavic–Indo-Iranian, Italic-Germanic-Celtic)

There are two clear **Greek-Armenian** lexical innovations in the Swadesh list:

1) reflexes of *$h_2oi̯u$-k^wid* replace PIE *ne* 'not';
2) reflexes of PIE *h_2e/onh_1-mo-* 'breath' replace PIE *$h_2u̯eh_1$-nt-* 'wind'.

One more Greek-Armenian match — *$ksē/oro$-* 'dry' — is unlikely to be a common Graeco-Armenian innovation for the basic meaning 'dry', since some varieties of Ancient Greek, above all the lects of Homer and Herodotus, preserve αὖος < PIE *$h_2s(o)us$-o-s as the basic word for 'dry'.

Only one lexical innovation can be uncontroversially postulated for the **Balto-Slavic–Indo-Iranian** clade: reflexes of PIE *$pleu$-* 'float, flow' replace PIE *$sneh_2$-* 'swim' in Slavic, Indian and Iranian (Baltic has unrelated *$peld$-*).

The following cases are potential innovations, because we cannot be sure what the Proto-Indo-European words for these Swadesh meanings were.

1) Old Indic, Proto-Slavic and Baltic (namely Old Prussian which is not in our database) have reflexes of *$kr̥s$-no-* for 'black'.
2) Proto-East-Baltic and Proto-Iranian have reflexes of *$k̑olH$-to-* 'cold'.
3) For 'green' we have reflexes of PIE *$ǵ^helh_3$-* in Slavic, East Baltic, Old Indic and Iranian. Other IE branches have derivatives of this root showing that it probably was the main word for 'green / yellow' already in Inner IE.
4) Reflexes of *$u̯olk̑$-o-* mean 'hair' in Slavic and Iranian, but 'sprout, twig' in Old Indic.
5) Slavic, Old Indic and Iranian have reflexes of *g^wrH-* 'mountain'; East Baltic reflexes of this word have a clearly secondary meaning 'forest'. If Greek βορέας 'north wind' (< *'wind from the mountains') is related, *g^wrH-* 'mountain' can be reconstructed for Inner IE.
6) Reflexes of PIE *$pont$-* mean 'road' in Old Indic, Iranian, Slavic and Baltic (namely Old Prussian). Meanings like 'sea', 'ford', 'bridge' in other branches can be derived from 'path (a trajectory)' rather than 'road (a landscape element)'.

7) Slavic, East Baltic, Old Indic and Iranian words for 'stone' go back to *$h_2eḱ$-men-. The Greek reflex ἄκμων means 'anvil, meteoric stone', which may point to the meaning 'stone' for this word already in Inner IE.

8) Finally, East Baltic and Old Indic have *$ǵ^weh_2$- as a preterit/aorist stem of 'to go'. This aorist stem is also attested in Greek, so we may deal here with an archaism rather than an innovation.

Italic, Germanic and Celtic have a number of exclusive matches in the Swadesh list, some of which may be innovations. However, only one word is common to all three branches: *p(e)isk-o- 'fish' replaces PIE *$d^hǵ^huH$- in Latin, Old Irish and Proto-Germanic (Brittonic has a Latin loanword).

The remaining matches tie together only two of the three branches.

1) Reflexes of *h_2e/ol- mean 'all' in Old Irish and Proto-Germanic. Proto-Brittonic *holl is cognate to Oscan *sullus* 'each, all, whole', so we have two conflicting matches here. The PIE word for 'all' was probably *peh_2-nt-, preserved in Greek and Tocharian.

2) There are two conflicting matches for 'breast': *b^hreus- (Old Irish, Proto-Brittonic, Proto-Germanic) vs. *pekt- (Old Irish, Latin).

3) The word *kaput- 'head' is common to Latin and Proto-Germanic. Celtic has isolated *k^wenno- 'head'. The PIE word for 'head' was either *$g^heb^hh_2$-l-, preserved in Greek and Tocharian A, or *$ḱerh_2$-, whose various derivatives mean 'head' and/or 'horn' in many branches.

4) The word *luH- 'louse' is common to Proto-Brittonic and Proto-Germanic. This is probably an innovation, because the Tocharian cognate means 'animal', and semantic development 'animal' > 'louse' is attested in other cases. However, we cannot know whether this innovation occurred at the Italo-Celto-Germanic stage or already in Inner IE.

5) Latin and Proto-Germanic have reflexes of *kols-o- 'neck'. The PIE word for 'neck' is hard to reconstruct.

6) The word for 'skin' in Latin and Proto-Germanic goes back to PIE *kuH-ti-. Another reflex of this word is Tocharian A *kāc* 'skin, hide' (apparently not the basic word for human skin).

Supplement references

Adams, Douglas Q. 2017. Tocharian. In *The Indo-European languages*, ed. Mate Kapović, 2nd ed., 452–475. London: Routledge.

Adams, Douglas Q., and James P. Mallory. 1997a. Baltic languages. In *Encyclopedia of Indo-European culture*, ed. James P. Mallory and Douglas Q. Adams, 46–50. London: Fitzroy Dearborn Publishers.

Adams, Douglas Q., and James P. Mallory. 1997b. Indo-Iranian languages. In *Encyclopedia of Indo-European culture*, ed. James P. Mallory and Douglas Q. Adams, 303–312. London: Fitzroy Dearborn Publishers.

Aikio, Ante. 2002. New and old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57: 9–57.

Anthony, David W., and Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1: 199–219. https://doi.org/10.1146/annurev-linguist-030514-124812.

Beekes, R. S. P. 2010. *Etymological dictionary of Greek*. Leiden Indo-European Etymological Dictionary Series 10/1–2. Leiden: Brill.

Bouckaert, Remco R., and Joseph Heled. 2014. DensiTree 2: seeing trees through the forest. *bioRxiv*: 012401. https://doi.org/10.1101/012401.

Bousquette, Joshua, and Joseph Salmons. 2017. Germanic. In *The Indo-European languages*, ed. Mate Kapović, 2nd ed., 387–420. London: Routledge.

Cantera, Alberto. 2001. Die Behandlung der idg. Lautfolge (C)RHC- im Iranischen. *Münchener Studien zur Sprachwissenschaft* 61: 7–27.

Chindina, L. A. 2001. Kulajskaja kul'tura [Kulay culture]. In *Narody i kul'tury Tomsko-Narymskogo Priob'ja: Materialy k enciklopedii Tomskoj oblasti*, ed. È. I. Chernyak, 78–81. Tomsk: TGU.

Clackson, James P. T. 2008. Classical Armenian. In *The Ancient Languages of Asia Minor*, ed. Roger D. Woodard, 124–144. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511486845.014.

Helimski, Eugen. 1982. *Drevnejšie vengersko-samodijskie jazykovye paralleli [Ancient linguistic parallels between Hungarian and Samoyed]*. Moscow: Nauka.

Helimski, Eugen. 1997. *Die matorische Sprache : Wörterverzeichnis, Grundzüge der Grammatik, Sprachgeschichte*. Studia Uralo-Altaica 41. Szeged: Department of Altaic Studies, University of Szeged.

Helimski, Eugen. 2005. The 13th Proto-Samoyedic vowel. In *Mikola-konferencia 2004*, ed. Beáta Wagner-Nagy, 27–39. Szeged: SzTE Department of Finnougristics.

Helmer, Daniel, Lionel Gourichon, Hervé Monchot, Joris Peters, and Maria Saña Segui. 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In *First steps of animal domestication: new archaeozoological approaches*, ed. J. D. Vigne, J. Peters, and D. Helmer, 86–95. London: Oxbow Books.

Hill, Eugen. 2012. Hidden sound laws in the inflectional morphology of Proto-Indo-European. In *The sound of Indo-european: phonetics, phonemics and morphophonemics*, ed. Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen, and Jens Elmegård Rasmussen, 169–207. Copenhagen Studies in Indo-European 4. Copenhagen: Museum Tusculanum, University of Copenhagen.

Hongo, Hitomi, Jessica Pearson, Banu Öksüz, and Gülçin Ilgezdi. 2009. The Process of Ungulate Domestication at Çayönü, Southeastern Turkey: A Multidisciplinary Approach focusing on *Bos* sp. and *Cervus elaphus*. *Anthropozoologica* 44: 63–78. https://doi.org/10.5252/az2009n1a3.

Illich-Svitych, Vladislav Markovich. 1967. Materialy k sravnitel'nomu slovarju nostratičeskix jazykov [Materials for a comparative dictionary of the Nostratic languages]. In *Etimologija 1965*, 321–373. Moscow: Nauka.

Illich-Svitych, Vladislav Markovich. 1976. *Opyt sravnenija nostratičeskix jazykov [An attempt at a comparison of the Nostratic languages]*. Edited by V. A. Dybo. Vol. 2: Sravnitel'nyj slovar' (l — ʒ́). Moscow: Nauka.

Ivanov, Vyacheslav V. 2001. Southern Anatolian and Northern Anatolian as separate Indo-European dialects and Anatolian as a late linguistic zone. In *Greater Anatolia and the Indo-Hittite language family: papers presented at a colloquium hosted by the University of Richmond, March 18-19. 2000*, ed. Robert Drews, 131–183. Journal of Indo-European Studies, Monograph Series 38. Washington, DC: Institute for the Study of Man.

Janhunen, Juha. 1977. *Samojedischer Wortschatz: gemeinsamojedische Etymologien*. Castrenianumin Toimitteita 17. Helsinki: Helsingin yliopisto: Suomalais-ugrilainen seura.

Kassian, Alexei S. 2001. Otraženie praslavjanskoj fonemy *y v staropskovskom dialekte razgovornika T. Fenne kak arxaizm [Reflexes of the Proto-Slavic phoneme *y in the Old Pskov dialect of Tonnies Fenne's phrasebook as a retention]. *Voprosy jazykoznanija* 4: 73–100.

Kassian, Alexei S. 2011a. Annotated Swadesh wordlists for the Hittite (Old Hittite) language (Anatolian group, Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Kassian, Alexei S. 2011b. Annotated Swadesh wordlists for the Greek group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Kassian, Alexei S., George Starostin, Anna Dybo, and Vasily Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.

Kassian, Alexei S., George S. Starostin, and Mikhail A. Zhivlov. 2015. Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43: 301–347.

Kassian, Alexei S., and Ilya Yakubovich. 2013. Anatolijskie jazyki [Anatolian languages]. In *Jazyki mira: Reliktovye indoevropejskie jazyki Perednej i Central'noj Azii [Languages of the World: Relict Indo-European languages of Western and Central Asia]*, ed. Yu. B. Koryakov and A. A. Kibrik, 15–26. Moscow: Academia.

Kiparsky, Valentin. 1963. *Russische historische Grammatik*. Vol. 1: Die Entwicklung des Lautsystems. 3 vols. Heidelberg: Carl Winter.

Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden Indo-European Etymological Dictionary Series. Leiden: Brill.

Kocharov, Petr. 2017. Annotated Swadesh wordlists for the Armenian group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Korn, Agnes. 2016. A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121: 401–434. https://doi.org/10.1515/if-2016-0021.

Korn, Agnes. 2019. Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics* 9: 239–281. https://doi.org/10.1075/jhl.17010.kor.

Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden Indo-European Etymological Dictionary Series 11. Leiden: Brill.

Kulikov, Leonid. 2017. Indo-Aryan. In *The Indo-European languages*, ed. Mate Kapović, 2nd ed., 214–262. London: Routledge.

Kümmel, Martin J. 2008. Rev. of: Penney, J. H. W. (ed.). Indo-European perspectives: studies in honour of Anna Morpurgo Davies. *Kratylos* 53: 25–35. https://doi.org/10.29091/kratylos/2008/1/3.

Kümmel, Martin J. 2016. *syá- im Indoiranischen: Zahlwort und Demonstrativum? In *Tavet tat satyam: studies in honor of Jared S. Klein on the occasion of his seventieth birthday*, ed. Andrew Miles Byrd, Jessica DeLisi, and Mark Wenthe, 179–190. Ann Arbor: Beech Stave Press.

Kuritsyna, Anna. 2017. Annotated Swadesh wordlists for the Tocharian group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Kushniarevich, Alena, Olga Utevska, Marina Chuhryaeva, Anastasia Agdzhoyan, Khadizhat Dibirova, Ingrida Uktveryte, Märt Möls, et al. 2015. Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLOS ONE* 10: 1–19. https://doi.org/10.1371/journal.pone.0135820.

Lubotsky, Alexander. 2001. The Indo-Iranian substratum. In *Early contacts between Uralic and Indo-European: linguistic and archaeological considerations.*, ed. Asko Parpola, Petteri Koskikallio, and Christian Carpelan, 301–317. Helsinki: Suomalais-Ugrilainen Seura.

Lubotsky, Alexander, and Sergei A. Starostin. 2003. Turkic and Chinese loan words in Tocharian. In *Language in time and space: a Festschrift for Werner Winter on the occasion of his 80th birthday*, ed. Brigitte L. M. Bauer and Georges-Jean Pinault, 257–269. Trends in Linguistics Studies and Monographs 144. Berlin: de Gruyter.

Lucht, Martina. 2007. Der Grundwortschatz des Altirischen. Inaugural-Dissertation zur Erlangung der Doktorwürde, Bonn: Rheinische Friedrich-Wilhelms-Universität.

Mac Mathúna, Liam. 1978. On the expression of 'Rain' and 'It is raining' in Irish". *Ériu* 29: 39–57.

Mallory, James P., and Douglas Q. Adams. 1997. Germanic languages. In *Encyclopedia of Indo-European culture*, ed. James P. Mallory and Douglas Q. Adams, 218–223. London: Fitzroy Dearborn.

Napolskikh, Vladimir V. 1997. *Vvedenie v istoricheskuju uralistiku [Introduction to historical Uralistics]*. Izhevsk: Udmurt Institute for History, Languages and Literature.

Olsen, Birgit Anette. 1999. *The noun in Biblical Armenian: origin and word formation: with special emphasis on the Indo-European heritage*. Tübingen: Mouton de Gruyter.

Orel, Vladimir E. 2003. *A handbook of Germanic etymology*. Leiden: Brill.

Peyrot, Michaël. 2013. *The Tocharian subjunctive: a study in syntax and verbal stem formation*. Brill's Studies in Indo-European Languages & Linguistics 8. Leiden: Brill.

Puhvel, Jaan. 2011. *Hittite etymological dictionary*. Vol. 8: Words beginning with PA. Berlin: Mouton de Gruyter.

Rastorgueva, Vera S., and Joy I. Edelman. 2000-. *Ètimologičeskij slovar' iranskix jazykov [Etymological dictionary of the Iranian languages]*. Moscow: Vostochnaya Literatura.

Ringe, Donald A. 2017. *From Proto-Indo-European to Proto-Germanic*. 2nd ed. A Linguistic History of English 1. Oxford: Oxford University Press.

Rix, Helmut, Martin J. Kümmel, Thomas Zehnder, Reiner Lipp, and Brigitte Schirmer. 2001. *LIV, Lexikon der indogermanischen Verben: die Wurzeln und ihre Primärstammbildungen*. 2nd ed. Wiesbaden: Dr. Ludwig Reichert.

Saenko, Mikhail. 2013. Rekonstrukcija praslavjanskogo spiska Svodeša [Reconstruction of the Proto-Slavic Swadesh wordlist]. *Journal of Language Relationship* 10: 139–148.

Saenko, Mikhail. 2015. Annotated Swadesh wordlists for the Romance group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Schrijver, Pieter. 1995. *Studies in British Celtic historical phonology*. Leiden Studies in Indo-European 5. Amsterdam: Rodopi.

Schumacher, Stefan. 2000. *The historical morphology of the Welsh verbal noun*. Maynooth Studies in Celtic Linguistics 4. Maynooth: The Department of Old Irish, National University of Ireland.

Sedov, V. V. 1995. *Slavjane v rannem Srednevekov'e [Slavs in the Early Middle Ages]*. Moscow: Institut arxeologii (Rossijskaja akademija nauk).

Sihler, Andrew L. 1995. *New comparative grammar of Greek and Latin*. 1st ed. New York: Oxford Univ. Press.

Sims-Williams, Nicholas. 1996. Eastern Iranian languages. *Encyclopædia Iranica*. London.

Sims-Williams, Nicholas. 2017. Iranian. In *The Indo-European languages*, ed. Mate Kapović, 2nd ed., 263–286. London: Routledge.

Sims-Williams, Patrick. 2017. Celtic. In *The Indo-European languages*, ed. Mate Kapović, 2nd ed., 352–386. London: Routledge.

Starostin, George S. 2011. Annotated Swadesh wordlists for the Albanian group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Starostin, George S. 2016. Annotated Swadesh wordlists for the Germanic group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Sussex, Roland, and Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486807.

Tedesco, Paul. 1921. Dialectologie der mitteliranischen Turfantexte. *Monde oriental* 15: 184–258.

Tischler, Johann. 2004. *Hethitisches etymologisches Glossar*. Vol. 3. Innsbrucker Beiträge zur Sprachwissenschaft 20. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Trofimov, Artem. 2016. Annotated Swadesh wordlists for the Indo-Aryan group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Trofimov, Artem, and Oleg Belyaev. 2014. Annotated Swadesh wordlists for the Iranian group (Indo-European family). In *Global lexicostatistical database*, ed. George S. Starostin. http://starling.rinet.ru/cgi-bin/main.cgi?root=new100.

Vine, Brent. 1999. Greek ῥίζα 'root' and 'Schwa Secundum.' In *UCLA Indo-European Studies*, ed. Vyacheslav V. Ivanov and Brent Vine, 1:5–30. Los Angeles, CA: University of California, Los Angeles, Program in Indo-European Studies.

Watkins, Calvert. 2008. Hittite. In *The Ancient languages of Asia Minor*, ed. Roger D. Woodard, 6–30. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511486845.005.

Windfuhr, Gernot. 2009. Dialectology and topics. In *The Iranian languages*, ed. Gernot Windfuhr, 5–42. Routledge Language Family Series. London: Routledge.

Witzel, Michael. 2003. Vedas and Upanisads. In *The Blackwell companion to Hinduism*, ed. Gavin Flood, 66–101. Oxford: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470998694.ch4.

Yakar, Jak. 2012. Anatolian chronology and terminology. In *The Oxford handbook of Ancient Anatolia*, ed. Gregory McMahon and Sharon Steadman, 1:56–93. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376142.013.0004.