Universiteit
Leiden
The Netherlands

Unde venisti? The Prehistory of Italic through its Loanword Lexicon

Wigman, A.M.

Citation
Wigman, A. M. (2023, November 1). *Unde venisti?: The Prehistory of Italic through its Loanword Lexicon*. Retrieved from https://hdl.handle.net/1887/3655644

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3655644

Note: To cite this publication please use the final published version (if applicable).

4 Distribution Analysis

4.1 Introduction to Distribution Analysis

There is a strong pattern amongst the Latin words of non-inherited origin. Of the 109 words from §2.2, 93 have comparanda in Celtic, Germanic, Greek, or a combination of the three. Of the 16 words without comparanda in one or more of these three branches, there are:

- 15 words isolated to Italic:
    - 3 words without certain comparanda (*asīlus*, *asinus*, *casa*) that represent recent loans due to their intervocalic *s*,
    - 5 words without certain comparanda (*faex*, *farciō*, *focus*, *pampinus*, *tabānus*) that represent loans due to their invalid root structure,
    - 7 words (*arbutus*, *cerrus*, *genesta*, *lābrusca*, *lepus*, *sambūcus*, *talpa*) with irregular alternations clearly attested within Latin and/or Romance, and
- 1 word (*faber*) with a comparandum only in Armenian, probably a Wanderwort from an ultimately Hurrian source.

There are additionally two additional words amongst the 93 whose comparanda in Celtic/Germanic/Greek are not fully secure. *Carpinus* may only have secure comparanda in Slavic and *avēna* in Baltic and Slavic.

The 93 words with Celtic/Germanic/Greek comparanda pattern as follows (Figure 4.1). Words in non-italics are from §2.2.2.1 and those in italics are from §2.2.2.2. Words in green have no comparanda outside of the constellation of Celtic/Germanic/Greek where they are listed. Four words are preceded by a question mark because some comparanda are uncertain;[522] not enough to greatly influence the results.

[522] On *carpinus* and *avēna*, see above. The Celtic comparandum for *raudus* is not as secure as the Germanic one. *Tilia* has matches in Greek and Celtic, possibly Germanic. *Columba* is quite likely not actually found in Greek.

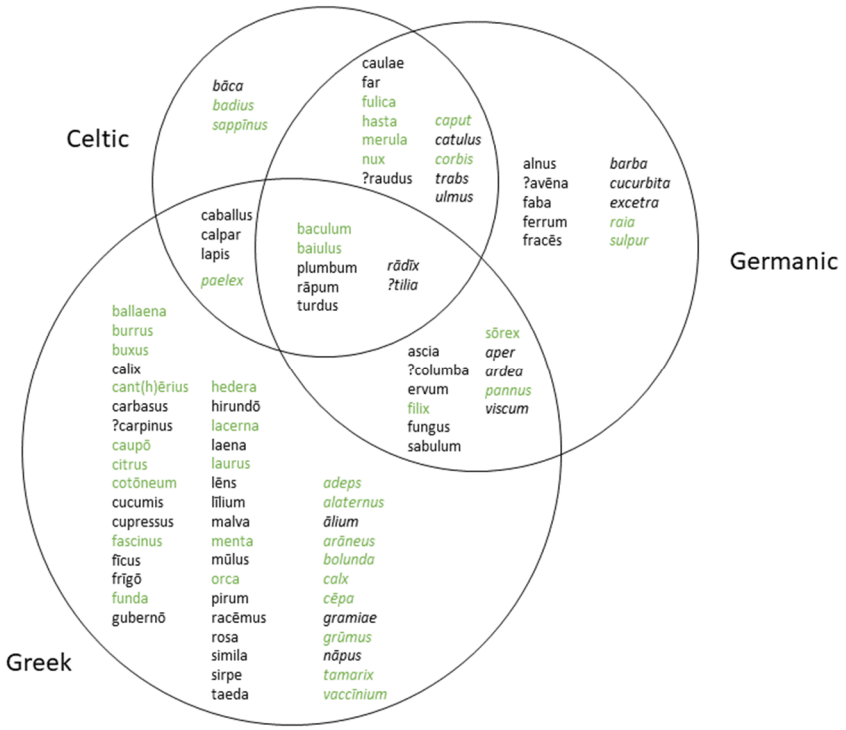

Figure 4.1 Latin words of loanword origin distributed by existence of comparanda in Celtic, Germanic, and Greek

The importance of Greek is immediately clear. 68 of 109 (62%) Latin words of non-inherited origin have Greek comparanda. For 46 of those 68 (68%), Celtic and Germanic do not attest comparanda. And 23 of those 46 (50%) are exclusive Latin-Greek isoglosses. Many of these must have been acquired in the Mediterranean. For example, 2 or 3 (*sirpe*, *taeda*, ?*lēns*) have independent comparanda in Berber; 6 (*cupressus*, *fīcus*, *laena*, *malva*, *rosa*, *simila*) in Semitic. The comparanda for *laena* and *simila* are well-attested in Semitic such that it might be the ultimate source of the Latin and Greek words. *Rosa* is also deeply entrenched in Semitic, but seems to have its ultimate source further East and cannot be the direct source of the Latin. *Cupressus*, *fīcus*, and *malva* have a much more limited distribution in Semitic. There is a good chance that they represent independent borrowings. Hebr. *gofer* is a hapax, but attests to the lexeme behind Gk. κυπάρισσος without the suffix. The Latin form has the suffix but cannot be a direct borrowing from Greek. This suggests a mode of transmission like in Figure 4.2.

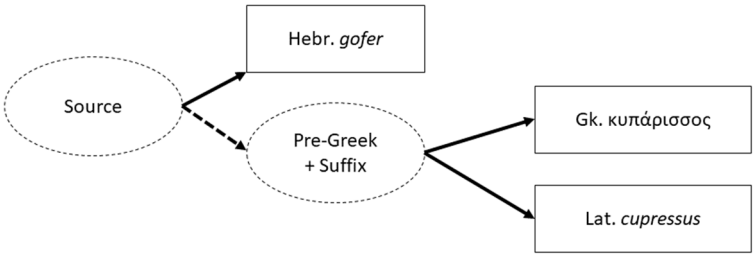

Figure 4.2 The transmission of Lat. *cupressus* and its comparanda

Hebr. *šiqmā* 'sycamore fig' is not as isolated, but occurs only in Hebrew and Aramaic (the latter likely being the source of Arabic *suqūm*, Klein 1987: 679). Lat. *fīcus* mechanically reconstructs to PIE **d^h*, but in light of Gk. τῦκον and Arm. *t^{c}uz*, there are two possibilities: 1) **t^h* can be reconstructed for Italic, Greek,[523] and Armenian, suggesting that the word was borrowed with the same initial consonant or 2) the word entered Latin with **θ*, which is also one of the Proto-Semitic sources of Hebr. *š*. In either case, the form borrowed into Latin (whether **θ* of **t^h*) entered Latin's treatment chain of the voiced aspirates, showing it was borrowed still into Proto-Italic.

On the other side of the spectrum are the 25 words for which there are Celtic and Germanic comparanda to the exclusion of Greek. 6 of these 25 (24%) are Latin-Celtic-Germanic isoglosses. This likely represents words borrowed in a part of Europe whose substrate languages did not affect Greek. The possibility of an Italo-Celtic subnode has interesting implications for this group, especially given the small number of Latin-Celtic isoglosses (*badius* and *sappīnus*, though *bāca* has only Berber additionally).

The other significant portion is at the center—the 7 words that have comparanda in Celtic, Germanic, and Greek. 2 of these 7 (29%) are Latin-Celtic-Germanic-Greek isoglosses. Others like *plumbum*, *rāpum*, and *turdus* are quite widespread. Their distributions are likely the result of different types of contact situations, as will be discussed below.

The picture that emerges based on the distribution of loanwords within these four daughter branches must be an oversimplification in many respects. Thus the following sections treat the stratification using further evidence. But a clear distribution is already apparent: the most recent stratum of loanwords in Latin was acquired in the Mediterranean. It is quantitatively the largest, and contains several of the words that cannot be reconstructed beyond Latin even. In contrast, the number of words with a pan-European distribution is much smaller. This suggests that these are the oldest loans,

[523] Recall that Att-Ion. σῦκον points to the presence of a glide that would neutralize the difference between PGk. **t* and **t^h* (s.v. *fīcus*).

borrowed at a time when the daughter languages were closer to one another followed by lexical replacement.

4.2 Partial Stratification Based on Phonology: The Most Recent Borrowings

Certain: *asinus, asīlus, casa, carbasus, ballaena*
Possible: *caballus, rosa, tamarix*, (*cāseus, gabata, nepeta, supparus*)

We know that PItal. **s* was rhotacized in Latin in the historical period, ca. 350 BCE (cf. Weiss 2020: 208). Thus Latin words with unrhotacized *s* not following a long vowel/diphthong must have entered Latin after rhotacism took place. Schrijver (1991: 252) points to *asinus* and *cāseum* to suggest that rural dialects that did not undergo rhotacism might be the source of some of these unrhotacized forms. Christol (1996) proposes a stricter version of this, where non-rhotacizing dialects preserved intervocalic *z*, sometimes borrowed back into Latin as *ss*. Geminate *s* is also the result of borrowed Gk. ζ, and it explains e.g. Plautine *nassus* otherwise *nāsus* 'nose' beside *nāris*. *Asinus*, *asīlus*, *casa*, and *carbasus* cannot be explained according to this stricter framework and would thus be post-rhotacism loans. Weiss (2020: 162) suggests that the *s* of *rosa* remained unrhotacized due to the dissimilatory effects of the *r*, like in *miser* 'wretched'. *Cāseus*, given its long vowel, might attest to a simplified geminate rather than a true unrhotacized *s*; its analysis is hampered by a lack of secure comparanda.

Vowel weakening began to occur sometime after 500 BCE. The earliest change was the weakening in non-initial syllables of *a* to *e*, and in open syllables further to *i* (cf. Weiss 2020: 130). *Ballaena* should have weakened to ***balleina* and been monophthongized to ***ballīna*, so it was borrowed after vowel weakening. The same could be the case for *caballus* and *tamarix*, but it cannot be ruled out that their medial vowels resisted weakening due to the *alacer rule* whereby a short vowel in a medial syllable is preserved if it contains the same vowel as that of the initial syllable and is separated by only one consonant (cf. Weiss 2020: 128-9). For the same reason, it is difficult to prove when *gabata* and *nepeta*, most likely loans from Greek, were borrowed. The unweakened *a* of *supparus* cannot be explained in this way, but there exists the possibility that it is a loan from Oscan.

Importantly, *asinus*, *asīlus*, and *casa* have no secure comparanda, but *ballaena* is a Latin-Greek isogloss. The Latin and Greek forms are very similar, but slight differences mean that the Latin word cannot be a direct borrowing from Greek. This suggests the existence of (an) unknown language(s) of the Mediterranean that served as an intermediary of Greek words into Latin or from which both Latin and Greek borrowed. *Ballaena* could be seen as preserving the αι diphthong of Gk. φάλλαινα, but it cannot be ruled out that the source word had this diphthong. In any case, this language must have been in existence at a very late date: after vowel weakening. *Carbasus* is somewhat more

complex in light of its other comparanda, but it is a crucial case. Its unrhotacized *s* places it amongst the most recent Latin loanwords. It is a close match with Gk. κάρπασος, but the alternation in voicing shows it, like *ballaena*, is not a direct loan. If *carbasus* and *ballaena* belong to this most recent stratum of loanwords into Latin, then probably so too do other Latin-Greek isoglosses that show similar types of alternations.

4.2.1 Latin-Greek Isoglosses with Recurring Irregular Alternations

4.2.1.1 Voicing and Devoicing

Like Lat. *carbasus* ~ Gk. κάρπασος, there is a group of Latin-Greek isoglosses that differ in the voicedness of their consonants.

Lat. *buxus* is identical to Gk. πύξος in form and meaning but for the irregular *b* ~ *p* alternation. It occurs also between Lat. *burrus* ~ Gk. πυρρός (as well as between the same forms in use as personal names). It may be present in Lat. *burgus* ~ Gk. πύργος, but there is a chance that the Latin form is a loan from Germanic (Biville I: 235-7).

A similar voicing alternation occurs between Lat. *gubernō* ~ Gk. κυβερνάω. Since Greek has variants like Cypriot *ku-me-re-na-i*, the closeness of the Latin form to the Greek form might mean that, regardless of the source of the Greek form, it was the ultimate source of the Latin, mediated by an unknown language. Such is also the case for Lat. *grūmus* ~ Gk. κρῶμαξ, κλῶμαξ with its *l* ~ *r* alternation.[524]

The opposite voicing correspondence occurs in Lat. *cotōneum* ~ Gk. κυδώνια and Lat. *taeda* ~ Gk. δαΐς. It may also occur in Lat. *citrus* ~ Gk. κέδρος, but here the devoicing may be regular.

4.2.1.2 Aspiration Alternations

Like Lat. *ballaena* ~ Gk. φάλλαινα, several Latin-Greek isoglosses attest to aspiration alternations. The inherited source of Gk. φ is PIE **b^h*, so one might propose that the pair *ballaena* : φάλλαινα represents an older loan, closer to the time at which PGk. **p^h* was developing from **b^h*. But we have already seen that the lack of vowel weakening in *ballaena* requires it to be quite a recent loan in Latin.

Otherwise, Greek generally has a form that looks like the reflex of a voiced aspirate against a Latin *tenuis*. Because this is the normal outcome in Latin loanwords from Greek (cf. Biville I: 142), these cases critically show other irregularities that demonstrate that they are not direct loans. The alternations include **d^h* ~ **t*: Lat. *menta* ~ Gk. μίνθη (with additional *e* ~ *i* alternation), Lat. *lēns* ~ Gk. λάθυρος (vocalism points to syllabic *n*, too early for borrowing), Lat. *cant*(*h*)*ērius* ~ Gk. κανθήλιος (with *l* ~ *r* alternation in

[524] Other cases of this correspondence are difficult to analyze due to the occurrence of unvoiced variants in Latin (see fn. 75). The opposite correspondence might occur between Lat. *conger* ~ Gk. γόγγρος, but the *c* might be the result of remodeling in Latin on other words beginning with *con-*.

what is otherwise looks to be the same suffix); **g^h* ~ **k*: Lat. *orca* ~ Gk. ὕρχη (with additional *o* ~ *u* alternation), Lat. *calx* : Gk. χάλιξ (with unexplained syncope); **b^h* ~ **p*: Lat. *adeps* ~ Gk. ἄλειφα(ρ) (with *d* ~ *l* alternation). Cf. also Lat. *sirpe* ~ Gk. σίλφιον (with additional *l* ~ *r* alternation), which has clear Berber comparanda. The Hesychian variant σέλπον attests to an aspiration alternation within Greek for this lexeme. This is also the case for Gk. κυπάρισσος, which has an aspirated variant preserved in κυφαρίσσινος.

4.2.2 Wider Implications: A Mediterranean Substrate

Thanks to the phonological features of *ballaena* and *carbasus* that require them to be recent loans in combination with the alternations they attest to with their Greek comparanda, we can conclude that a number of Latin-Greek irregular isoglosses can potentially be dated to quite recently. Nor is it surprising that Mediterranean contact situations should have been the most recent; both Latin and Greek ended up being spoken in the Mediterranean.

But while some loanwords seem to have entered Latin in the historical era, others must have entered at a Proto-Italic date. This includes *fīcus*, as mentioned above, but also probably the Latin-Greek isogloss *fascinus* ~ βάσκανος, which seems to have entered Italic either as **b^h* or an intermediate stage of this phoneme's development to attested *f* (see further under §4.3.2.1). Given this time span, it is important to consider the question of how many different languages Italic was in contact with in the Mediterranean and whether any of these was actually exclusive to the Mediterranean region. In short, is there really a Mediterranean substrate in Latin?

4.2.2.1 Other Words with a Mediterranean Distribution

A problem with the search for a Mediterranean substrate is (and has been) the difficulty in delineating which areas count as Mediterranean. Here I will examine the distribution of the words that have comparanda in Greek to the exclusion of Celtic and Germanic. This entails, including the cases discussed above:

> Several Latin-Greek isoglosses: *adeps*, *alaternus*, *arāneus*, *ballaena*, *bolunda*, *burrus*, *buxus*, *calx*, *cant*(*h*)*ērius*, *caupō*, *cēpa*, *citrus*, *cotōneum*, *fascinus*, *funda*, *grūmus*, *hedera*, *lacerna*, *laurus*, *menta*, *orca*, *tamarix*, *vaccīnium*
>
> Some Latin-Greek-Berber isoglosses: *sirpe*, ?*ālium*, ?*lēns*, ?*taeda*
>
> Some Latin-Greek-Semitic isoglosses: *cupressus*, *laena*, *malva*, *simila*
>
> A Latin-Greek-Armenian isogloss: *nāpus*
>
> A Latin-Greek-Armenian-Semitic isogloss: *fīcus*
>
> A Latin Greek-Albanian isogloss: *hirundō*
>
> And word with comparanda in Latin, Greek, Egyptian, perhaps Anatolian:

līlium.

Without any assumptions about time-depth, these distributional clusters are all arguably geographically Mediterranean. The only exception is Armenian, but it attests to other loanwords with a Mediterranean distribution (e.g. Arm. *xstor*, Gk. σκόροδον, Alb. *hurdhë* 'garlic', Clackson 2017: 1123).

There are also words with distributions whose relationship to the Mediterranean is not immediately clear.

4.2.2.2 Balkan Connections

The Latin-Greek-Albanian isogloss *hirundō* indicates a Mediterranean loanword with attestation in Albanian, which is not surprising given the geographic position of the Balkan Peninsula. There are however several words with an otherwise Mediterranean distribution that are also attested in Slavic. These are:

PSlav. **k/grъm-* ~ Lat. *grãmiae* ~ Gk. γλάμων

PSlav. **tyky-* ~ Lat. *cucumis* ~ Gk. σικύα, Hsch. κύκυον, σεκούα ~ Arm. *sex*

PSlav. **mъskъ-* ~ Lat. *mūlus* ~ Gk. μύκλος, Hsch. μυχλός, μύσκλοι ~ Alb. *mushk*

Since the migration of Slavic peoples into the Balkan Peninsula occurred in the sixth to seventh centuries CE (cf. Kobyliński 2005: 532-6), these words either entered Slavic after its speakers appeared in Balkans (and thus may still represent words from a Mediterranean language) or they attest to a different, earlier contact situation. The variation within Slavic of (roughly reconstructed) **k/grъm-* shows that it entered after the Common Slavic era and is thus an example of the former case. The *l* ~ *r* alternation between Lat. *grãmiae* ~ Gk. γλάμων is well-attested in other Mediterranean words (see §4.2.2.4.1). The same could be true for **tyky-*, whose comparanda are otherwise distributed similarly to *nāpus* and *fīcus*. It is more difficult to decide for **mъskъ-*. It seems to attest to the phenomenon of *SK* metathesis that otherwise sees a broader European distribution (cf. Lat. *ascia* ~ Gk. ἀξίνη ~ PGm. **akwes(ī)-*). Its status as a Balkan word might be artificially bolstered by the Albanian comparandum; it is likely a loan from Slavic. On the other hand, both variants are present in Greek, such that Slavic might have borrowed its word after arrival in the Balkans. Since donkeys were introduced to Italy and Greece from the East in the first millennium BCE (EIEC 33-4, Todd et al. 2022), it is quite unlikely that **mъskъ-* represents a loan alongside Italic and Greek from a time before the latter languages entered the Mediterranean region.

Attestation in Slavic therefore does not seem to rule out the possibility of a lexeme's origin in a substrate language of the Mediterranean when there are other indications that this is the case. Nor does it seem possible to extract from the Latin data a group of specifically Balkan origin.

Alessio (1946a) thought he found traces of a Mediterranean substrate in Baltic. Most of his evidence is almost exclusively toponymic, relying on the ability to see the same roots between e.g. hydronyms in the East and the West from which potential substrate lexemes can be distilled and assigned an aquatic/marine meaning. Amongst the lexical correspondences he does mention, three deserve mention:[525]

> Lith. *báltas*, Latv. *baĩts* 'white'[526] ~ PSlav. **bòlto-* 'swamp' ~ Alb. *baltë* 'mud, swamp', MoGk. βάλτος 'swamp' ~ Rom. *baltă* 'puddle, pool; swamp', dial. It. e.g. Lucchese *paltenna* 'mud'
>
> Lith. *korỹs*, Latv. *kâre* 'honeycomb' ~ Lat. *cēra* 'wax' ~ Gk. κηρός 'beeswax', κήρινθος 'beebread'
>
> Lith. *bríedis* 'elk, deer', Latv. *briêdis* 'elk' ~ Alb. *bri* 'horn, rack' ~ Hsch. βρένδον 'deer', Gk. βρέντιον (given as Messapic) 'stag's head' ~ PGm. **brinda*(*n*)- 'elk'

To these can be added:

> PSlav. **grabrъ* ~ Lat. *carpinus* ~ ?Gk. γράβιον ~ ?PBalt. **ske*/*irp-*
>
> Gk. κυβερνάω ~ Lat. *gubernāre* ~ ?Lith. *kum̃bryti* ~ ?OCS *krъmiti*

The group with Lith. *báltas* has been suggested to be an Illyrian substrate word (Brüch 1916, Derksen 2014 s.v. *baltas*). But given its late appearance outside Balto-Slavic (i.e. Romance and Modern Greek), the non-Balto-Slavic forms may all be early loans from Slavic (cf. Demiraj 1997: 87-8). As mentioned for Gk. κυβερνάω ~ Lat. *gubernāre*, the Balto-Slavic connections made by e.g. Boisacq (1916: 527-8) and Machek (1955: 61-4) may be the result of chance resemblance. The Baltic and even Greek comparanda for Lat. *carpinus* are not very good, and may simply be unrelated. Instead, it might attest to an Italo-Slavic isogloss and it is thus difficult to determine when this word was borrowed.

That leaves the groups of cervid and apicultural words. For the former group, Kroonen (2013: 77) separates the Germanic word as a coincidental lookalike with an internal Germanic etymology and Demiraj (1997: 110-11) notes formal difficulties with directly comparing the Albanian form. Derksen (2014 s.v. *briedis*) calls the relationship between the Baltic forms and the rest of the comparanda "obscure" but does not reject it. For the

[525] Some of his other examples include 1) Lat. *lāma* 'marsh' ~ Lith. *lomà* 'hollow, valley, plot', Latv. *lãma* 'hollow in a field or meadow, pool, pit' ~ PSlav. **lamъ* 'hollow, bend' ~ Gk. λάμια 'chasms'. But the Latin form is probably not related after all (Schrijver 1991: 142, Derksen 2014 s.v. *loma*). Nor can the Greek word be compared with any semantic certainty, removing the Mediterranean comparanda. 2) Latv. *ẽrms* 'monkey' ~ Etr. ăριμος 'monkey'. The Etruscan word is given by Strabo. The island today called Ischia was called by Virgil *Inarime* and *Arime* and was called by the Greeks Πιθηκοῦσαι, seemingly derived from πίθηκος 'ape'. No monkeys have lived on the island in recorded history (suggesting some sort of folk etymological interference), and the Etruscan word is too poorly attested to use.

[526] The semantic change that must be assumed to connect 'white' and 'swamp' is variously defended and viewed with suspicion.

latter group, Lat. *cēra* cannot be ruled out as loan from Gk. κηρός, but the suffix of κήρινθος proves its antiquity in Greece. It seems very unlikely that Baltic would ever have been in contact with the Mediterranean, at least in any direct linguistic way. Rather than a Mediterranean stratum in Baltic, an alternative explanation for these would be that they attest to a separate contact situation.[527] There are several cases of non-IE words that are attested in Baltic and Greek alongside others: Gk. λύγξ 'lynx' ~ PGm. **luhsa-* 'lynx' ~ Arm. *lusanun* 'lynx; hyena; marten' ~ Lith. *lūšis*, dial. *lųnšis* 'lynx'; Gk. θρύον 'reed, rush' ~ OCS *trъstь* 'reed, cane' ~ Lith. *triušìs* 'reed'; Gk. κρόμμυον 'onion' ~ PGm. **hramusan-* 'ramsons' ~ MIr. *crem* etc. 'wild garlic' ~ Ru. *čeremšá*, etc 'wild garlic' ~ Lith. *kermušẽ*, dial. *kermušà* 'tip or a drill or flail, wild garlic'. These distributions do not seem convincingly Mediterranean and Alessio's label seems to be the result of overstretching the term "Mediterranean" as the only substrate language of Europe.

4.2.2.3 Indo-Iranian Connections

Pisani (1938a) had been inspired by recent excavations of the Indus Valley Civilization to look for cultural connections stretching from the Mediterranean to India. He pointed to what he saw as cultural and religious similarities (like a Mother Goddess figure) in all of these areas in the Bronze Age. He only provided two lexical correspondences: Gk. ἐρέβινθος 'chickpea' : Skt. *aravinda-* 'lotus' and Gk. κάρπασος[528] 'fine flax' : Skt. *karpāsa-* 'cotton'.

Gk. ἐρέβινθος has many more irregular comparanda across Europe (Lat. *ervum*, PGm. **arwīt-*, Arm. *aṙowoyt*). While the appurtenance of the Sanskrit form Pisani mentions is disputed, there do indeed seem to be comparanda in Iranian languages. This word for an agricultural founder crop attests to an older contact situation with early farmers rather than a Mediterranean substrate language stretching to India. On the opposite side of the spectrum, Gk. κάρπασος is at most a Wanderwort with its origins in the East, having entered Greek after the loss of inherited intervocalic **s* and making its way to Latin (at an even later date) via some intermediary language of the Mediterranean. A similar situation, though it may well have occurred earlier, holds for Lat. *rosa* ~ Gk. ῥόδον.

There are a few other irregular Latin-Greek isoglosses (or near isoglosses) that have potential matches in Indo-Iranian however. These include:

Lat. *calix* ~ Gk. κύλιξ ~ ?Skt. *kaláśa-*

Lat. *frīgō* ~ Gk. φρῡ́γω ~ ?PIr. **bra*(*i*)*ǰ-*, Skt. *bhrajj-*

Lat. *pirum* ~ Gk. ἄπιον ~ ?Shina *pisō* ~ ?Burushaski *pheṣo* ~ ?Khinalug *bzi*

Lat. *racēmus* ~ Gk. ῥᾱ́ξ, ῥώξ ~ ??Alb. *rrush* ~ ?PIr. **raza-*

In all cases, the appurtenance of the Indo-Iranian material is uncertain. Formally and

[527] Contact between populations involved in the trade of Baltic amber would be plausible.
[528] He gives κάρβασος, but this form is not actually attested.

semantically, the case for Skt. *kaláśa-* is strongest, as it points to a reconstruction with *a*-vocalism like Latin. The Indo-Iranian comparanda of *frīgō* and ῥῥῡ́γω are semantically compelling, though not all connect them (cf. LIV2 s.v. **bʰreǵ-*; DV 212, 243). The appurtenance of all the pear words is difficult to evaluate. While PIr. **raza-* can reconstruct to **(H)reǵ-*, formally aberrant **leǵ(ʰ)-* is alternatively possible. In no case is an extended Mediterranean substrate the only conceivable explanation however. As for *ervum*, some distributions might attest to an older contact situation while, like *carbasus*, others may represent Wanderwörter. And others still may be due to chance resemblance.

4.2.2.4 Recurring Features

With some of the confounding variables removed from the dataset, we can search the list of words with a confirmed Mediterranean distribution for recurring irregular correspondences. In addition to the voicing and aspiration alternations mentioned under §4.2.1, three others appear: *l* ~ *r*, *e* ~ *i*, and *i* ~ *u*.

4.2.2.4.1 L ~ R Alternation

There are 10 instances of an *l* ~ *r* alternation. For 6 of these, the lexeme can be argued to have a Mediterranean distribution.[529] In all of these 6, the alternation exists between Latin and Greek and there is no secure Semitic comparandum. In the first 5 cases, Lat. *r* corresponds to Gk. *l*:

> Lat. *cant*(*h*)*ērius* ~ Gk. κανθήλια
>
> Lat. *grūmus* ~ Gk. κλῶμαξ, Hsch. κρῶμαξ (here the alternation also exists within Greek)
>
> Lat. *hirundō* ~ Gk. χελῑδών ~ Alb. *dallëndyshe*
>
> Lat. *sirpe* ~ Gk. σίλφιον, Hsch. σέλπον ~ Berber *azlaf*, *aselbu*, etc.
>
> Lat. *grãmiae* ~ Gk. γλάμων (as detailed above, Slavic comparanda were likely borrowed late)

In one case, the correlation of *r* to *l* is opposite:

> Gk. λείριον ~ Lat. *līlium*, Hitt. *alēl*, Copt. *hlēli*, etc.

There is one further case where an *l* ~ *r* alternation exists between Latin and Romance descendants:

> PRom. **darbo-* ~ Lat. *talpa*

Lat. *paelex* and Gk. παλλακή agree on *l* against OIr. *airech*. But this likely belongs to a different contact phenomenon given the pattern so far and since Celtic does not clearly belong to the Mediterranean region. The same is certainly true for Lat. *filix* ~ Gk.

[529]For *l*(*l*)/*r*(*r*) as a Mediterranean alternation, cf. Bertoldi (1942: 183-4, fn. 1), Hubschmid (1953: 72), more generally as non-IE in Alessio (1944a: 124).

βλῆχνον, βλῆχρον ~ PGm. **brekna(n)-*. In the group of widely distributed pigeon words (including Lat. *columba*), a variant with *r* is present only in Egyptian. Theoretically this could be due to this word's transmission to Egyptian by a language of the Mediterranean, but this is difficult to prove. In any case, 6 or 7 out of 10 cases of an *l* ~ *r* alternation occur in lemmata with a Mediterranean distribution.

4.2.2.4.2 E ~ I Alternation

There are additionally several cases of an *e* ~ *i* alternation between comparanda in the dataset. In 6 cases, the alternation occurs between branches, with 5 of these cases being clearly restricted to the Mediterranean.[530] In 3 of these 5, Latin has *e* for Gk. *i*:

Lat. *cupressus* ~ Gk. κυπάρισσος (~ Hebr. *gofer*)

Lat. *menta* ~ Gk. μίνθη[531]

Lat. *hedera* ~ Gk. κιθάρα

In 2 cases, the distribution is reversed:

Lat. *hirundō* ~ Gk. χελῑδών (~ Alb. *dallëndyshe*)[532]

Lat. *citrus* ~ Gk. κέδρος

The non-Mediterranean case is Lat. *avēna* ~ PSlav. **ovьsъ* ~ PBalt. **(a)viža*, whose additional Uralic and potential Germanic comparanda point to a different contact situation.

There is an additional case of an *e* ~ *i* alternation within Latin (*genesta*/*genista*) and two within Greek for lexemes that have a Latin relative:

Gk. σίλφιον, Hsch. σέλπον ~ Lat. *sirpe* ~ Berber *azlaf*, *aselbu*, etc.

Gk. σικύα, Hsch. σεκούα, κύκυον ~ Lat. *cucumis* ~ Arm. *sex* ~ PSlav. **tyky-*

If these inner-branch alternations are included in the count, then there are 9 instances of an *e* ~ *i* alternation in the data, 8 of which occur with a Mediterranean distribution.[533] Some of these cases also include an *l* ~ *r* alternation and the alternations in voicedness and aspiration detailed above.

4.2.2.4.3 I ~ U Alternation

The set including Gk. κύκυον above also attests to *u*-vocalism. Another case of *i* ~ *u*

---

[530] *e*/*i* as a Mediterranean feature with different explanations also in e.g. Bertoldi (1939b: 89), Battisti (1959: 154-7).

[531] Romance languages also attest to another case like this: PRom. **plenta-* 'clod of earth' ~ Gk. πλίνθος 'brick' (cf. Alessio 1944a: 139-41).

[532] Its root vocalism, neither **e* nor **i*, perhaps suggests that Albanian was less directly involved.

[533] It seems very unclear whether Lat. *filix*/*felix* < **bʰel-ik-* ~ Gk. βλῆχνον, βλῆχρον < **bʰl-ēgʰ-r/n-* ~ PGm. **brekna(n)-* < **bʰr-eg-n-* is an example of this same type of alternation.

vocalic alternation is that of Lat. *supparus* ~ Gk. σίφαρος, σείφαρος. The Latin word may be a loan from Oscan, which makes it difficult to analyze the other alternations (geminate *pp* for Gk. φ, unweakened *a*) as original. But even if this is true, the *i* ~ *u* alternation lacks an explanation and makes it likely that the Oscan word was indirectly borrowed. This is reminiscent of two words with a clear *ī*/*ū* alternation: Lat. *fīcus* ~ Gk. τῦκον, σῦκον ~ Arm. *tʿuz* and Lat. *frīgō* ~ Gk. φρῡ́γω (Indo-Iranian comparanda, as detailed above, may not have much bearing on the words' treatment in the Mediterranean). Such *ĩ* ~ *ũ* alternations have been considered characteristic of the Mediterranean substrate (e.g. Bertoldi 1948: 70, Hubschmid 1953: 28, Alessio 1955: 375, 537-40, Battisti 1959: 155) and the examples are indeed restricted to the Mediterranean region. In fact, Lat. *hirundō* ~ Gk. χελῑδών attests to a *u* ~ *ī* alternation in its second syllable alongside *e* ~ *i* and *l* ~ *r* alternations.

4.2.2.5 A Definitively Mediterranean Substrate

A remarkable pattern emerges wherein the words with a Mediterranean distribution attest to a set of irregular alternations that are also by and large restricted to a Mediterranean distribution. The alternations overlap, such that a single lexeme can attest to more than one. Because they co-occur in the same word, it is highly likely that they are the result of the same language or at least group of closely related languages. The co-occurrence of the Mediterranean *e* ~ *i*, *i* ~ *u*, and *l* ~ *r* alternations in the same words as alternations in voicing and aspiration means that this Mediterranean substrate language was also responsible for at least some cases of this alternation (Figure 4.3).

The feature of voicing and aspiration alternation is more widespread than the Mediterranean. But rather than concluding from this that the Mediterranean substrate language extended far into Europe, it seems most likely that such alternations can be the result of different contact situations.

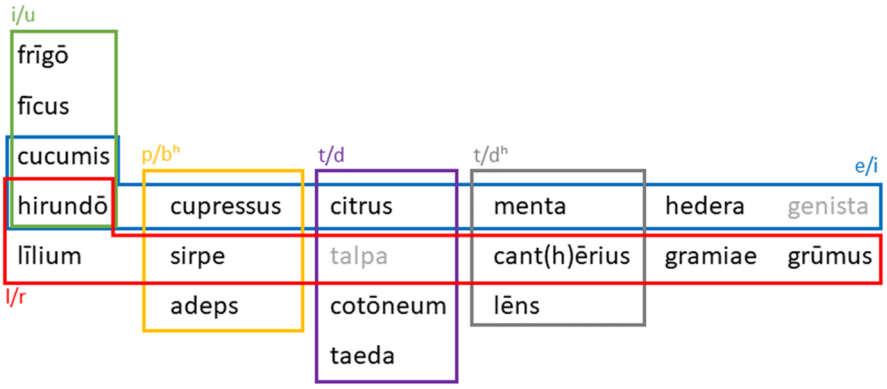

Figure 4.3 Overlapping irregular alternations between words with a Mediterranean distribution. (Words in gray are restricted to Italic)

It is interesting to note that the Mediterranean group contains *fīcus*, a word which must have entered at a Proto-Italic date. Thus it is likely that this language was present on the Italian peninsula and represents one of the earliest Mediterranean contact situations in the history of Latin. Though unclear in Figure 4.3, Lat. *talpa* alongside PRom. **darbo*- also attests to a *b* ~ *p* alternation that is reminiscent of the one represented by Lat. *carbasus* ~ Gk. κάρπασος. It is therefore possible the language persisted in the Mediterranean to a recent date. However, it should not be ruled out that there were other languages in the Mediterranean as well. Especially in light of cases that exhibit other alternations than the ones highlighted here, the language responsible for the *l* ~ *r*, *e* ~ *i*, and *i* ~ *u* alternations represents *at least* one distinct Mediterranean substrate language.

That some irregular alternations appearing in words restricted to a Mediterranean distribution are themselves more widespread brings up an important consideration. Claiming that words originated from (a) Mediterranean substrate language(s) based on their distribution alone is risky. What if the reason that the lexemes are restricted the Mediterranean is due to the fact that their *signifiés* are found only in the Mediterranean? It is plausible that, within one language family, languages in different ecological zones will have portions of mutually exclusive vocabulary based on the presence of something in one language's environment but its absence in that of the other. As concerns the Latin substrate vocabulary with a Mediterranean distribution of attestation, we see this for several plant names. The geographic range of the plants referred to by *alaternus*, *buxus*, *cupressus*, *fīcus* (and possibly *tamarix*[534]) themselves have a Mediterranean distribution (Figures 4.4-7). The species to which *cucumis* refers are likely of Asian origin (Sebastian, Schaefer, Telford & Renner 2010), and early evidence of them in the West comes from Egypt, Mesopotamia, and Greece (Zohary, Hopf, and Weiss 2012: 154-5) meaning their distribution as concerns Europe is Mediterranean.

[534] Heywood et al. (2007) give a distribution for the genus *Tamarix* that stretches into central Northern Europe, but most individual species do not seem to range beyond Southeast Europe.

Figure 4.4 Range of *Rhamnus alaternus* (Based on Bolòs & Vigo 1984-2001)

Figure 4.5 Range of *Ficus carica* (Based on Zohary, Hopf & Weiss 2012: 129)

Figure 4.6 Range of *Buxus sempervirens* (https://commons.wikimedia.org/wiki/File:Buxus_sempervirens_range.svg with refs.)

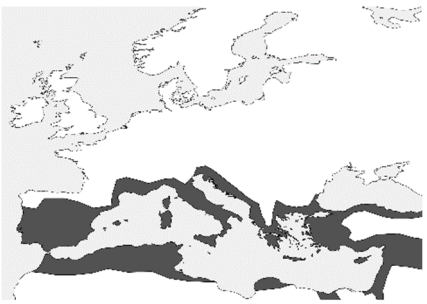

Figure 4.7 Range of *Cupressus sempervirens* (https://commons.wikimedia.org/wiki/File:Cupressus_sempervirens_range.svg with refs.)

However, several other plant words with a Mediterranean distribution, including some included in Figure 4.3, indeed refer to plants whose distribution is not limited to the Mediterranean. This includes *hedera*, *malva*, *menta*, and *vaccīnium*. In combination with the overlapping Mediterranean alternations (for *hedera* and *menta*), this suggests that something more than the geographic range of *signifiés* is producing the pattern. Most likely not in all cases, but at least in many, it instead points to a true linguistic boundary.

Given that *cupressus* and *hedera*, Latin words that seem to contain Pre-Greek suffixes, belong to the distinct Mediterranean substrate language isolated above, it seems to have been responsible for the indirect transmission of some Greek words into Latin.

4.2.2.6 Mediterranean Languages with a Wider Distribution

There is a small number of curious cases where non-Indo-European Mediterranean languages attest to comparanda that show a wider distribution. This is the case for Semitic and Sumerian. These must represent older contact situations, quite probably in

the way of Wanderwörter. This is relatively certain for Lat. *ferrum* and its Germanic, Kartvelian, and (probably) NE Caucasian comparanda. Along with several Semitic forms, these words have their ultimate source in an Anatolian language. It is in Anatolia that iron production technology developed, and it spread after the split of the Indo-European languages (cf. Thorsø & Wigman et al. 2023: 120). The group Lat. *ascia* ~ Gk. ἀξίνη ~ PGm. **akwes(ī)-* has comparanda in Semitic and Sumerian, suggesting a Wanderwort from the East. The Germanic form almost certainly precludes a borrowing from a Mediterranean contact situation.[535] But alternations within Akkadian (whence the other Semitic forms seem to be loans) indicate that it is not native there. All forms could be Wanderwörter with an ultimate source in Sumerian, but it cannot necessarily be ruled out that the Sumerian word is also a borrowing. A similar case with strong implications is that of Lat. *raudus* ~ PGm. **arut-* (and possibly PCelt. **rutu-*). They represent a case of non-IE *a*-prefixation, strongly suggesting that they entered from a European substrate source, but their similarity to OSum. *aruda* raises questions that are difficult to answer. Does this hint at a relationship between the language of the *a*-prefixes and Sumerian (cf. Schrijver 2018: 361-3), or is *aruda* a Wanderwort in Sumerian?

The case of Arm. *tʿuz* with its Latin (*fīcus*), Greek, and Semitic comparanda indicate that Armenian was in contact with some of the same Mediterranean languages as Latin and Greek. But several other cases of more widespread lexemes with attestation in Armenian show that it participated in some older contact situations involving Latin as well. Such situations as these will now be examined, primarily via their distributions of attestation amongst the IE daughter languages.

4.3 Further Stratification Using Distribution

4.3.1 Potentially Recent Borrowings

4.3.1.1 Words Exclusive to Latin and Romance

There are 5 words restricted to Latin (i.e. without secure comparanda) but which are nevertheless unlikely to be inherited due to their invalid root structure: *faex*, *farciō*, *focus*, *pampinus*, *tabānus*. The antiquity of these words in Latin is difficult to estimate given their lack of comparanda. They may represent words borrowed from a language spoken exclusively on the Italian peninsula, but it cannot be ruled out that their comparanda in other languages—and therefore true distribution—have been lost.

On the other hand, the 7 words restricted to Italic with irregularities recorded by Romance descendants (or alternations within Latin) most likely represent a recent

[535] Alessio (1941b: 204, 1946a: 143) had additionally used Ger. *Mohn* 'poppy', Gk. μήκων 'poppy', etc. to suggest that Germanic indeed has words from a Mediterranean substrate. He entertains part of the proposal by Ribezzo (1934a) that his would have occurred via contacts in the Alps with Rhaetic. Kroonen (2013: 371) however shows that all attestations of this word (which includes Baltic, Slavic, and Albanian) can be reconstructed to **méh₂k-on-*, an IE-looking root albeit only attested in Europe. This seems to be another case of the over-Mediterraneanization of European linguistic features.

stratum of loans, borrowed at least after the dissolution of Proto-Italic, if not after the development of the Latin branch. These include *arbutus*, *cerrus*, *genesta*, *lābrusca*, *lepus*, *sambūcus*, and *talpa*. It is *genesta* and *sambūcus* which show alternations within Latin, making them likely the oldest borrowings of the group. *Genesta* (var. *genista*) attests to an *e* ~ *i* alternation, quite likely the same one identified for the Mediterranean substrate (§4.2.2.4.2). Similar is likely true for the *l* ~ *r* alternation between Lat. *talpa* and PRom. **darbo*- (§4.2.2.4.1). They may thus represent examples where comparanda in other Mediterranean languages have been lost.

Sambūcus attests to a *b* ~ *mb* alternation within Latin. This otherwise appears in the Romance descendants of Lat. *lābrusca*. It is unclear if this is related to the *b* ~ *m* alternation between the Latin and Romance forms of *arbutus*, but it is interesting that Lat. *plumbum* attests to a nasal that is rather difficult to account for. While lead is present in the Mediterranean from quite an early date, the wide variation in the comparanda of Lat. *plumbum* makes it likely to be a relatively late Wanderwort, perhaps introduced along with the introduction of lead-alloyed copper that reaches Southern Europe around 1000 BCE (Thorsø & Wigman et al. 2023: 119). The irregular appearance of the nasal before *b* in *plumbum*, *sambūcus*, and *lābrusca* may thus be a phenomenon peculiar to the Italian peninsula, representing a feature of a language spoken there.

4.3.1.2 Gemination

Irregular Romance reflexes contain geminates not matched in Latin for *bāca*, *baculum*, *lapis*, and *lepus*. That a substrate language of the Italian peninsula might be responsible for the presence of geminates is suggested by the irregular alternation within Latin between *sappīnus* and *sabīna* (and perhaps the geminates present in Lat. *cerrus* and *vaccīnium*). The presence of both variants suggests, as above, a recent date of entry. However for *bāca*, *baculum*, and *lapis*, where Romance descendants attest to a geminate not recorded in Latin, comparanda exist outside of Italic (Greek, Germanic, and Celtic). Given that a similar pattern occurs between Lat. *lepus* ~ PRom. **lapparo*-, where non-Italic comparanda do not exist, it seems possible that the geminate forms may represent secondary reborrowing after the phonemicization of gemination in Latin and after the expansion of Latin back out of the peninsula (resulting in the Romance reflexes). This chronological differentiation may explain the pair *sappīnus* ~ *sabīnus* (and perhaps *bāca* ~ *vaccīnium*, if they are ultimately related).

4.3.2 Earlier Strata

4.3.2.1 The Oldest Loans in Italy

As mentioned above in §4.2.2, several lexemes from the corpus can be argued to have been borrowed at a time prior to or at the latest during the Proto-Italic developments regarding the voiced aspirates. The case is easiest to demonstrate when the Latin result differs in place of articulation from the reconstruction suggested by the comparanda. While it could be argued that a correspondence like Lat. *f* ~ PGm. **b* (in words with

other indications of being non-native) represents different nativizations of a foreign /β/ or /v/, that Lat. *f* ~ Gk. φ could represent words that entered with /f/ (especially when they are likely to be recent borrowings), and that Lat. *h* ~ Gk. χ could represent nativizations of a foreign /x/, similar explanations are more difficult when Lat. *f* points to the reconstruction of **d^h* or **g^{wh}*.

That loanwords entered such a stage in the history of Latin is not overly surprising for the words with a non-Mediterranean distribution. Such a case can be made for the following words:

In Initial Position

Likely

fracēs < **d^hrak-*: Guaranteed by PGm., PSlav. PBalt. onset **dr-*

fungus < **g^{wh}ong-*: Suggested by Gk. σφ/πόγγος, PGm. **swamb*/*ppan-*

Plausibly

faba < **b^hab-*: Suggested by PGm. *baunō-*, PBSl. **ba*/*ob-*

ferrum < **b^herso-*: Suggested by PGm. **brasa-*

filix/*felix* < **b^helik-*: Suggested by Gk. βλῆχνον, PGm. **brekna*(*n*)-

fulica < **b^hulVk-*: Suggested by PGm. **balikōn-*

hasta < **g^hast-*: Suggested by PCelt. **gasdo-*, **gazdo-*, PGm. **gazda-*

In Medial Position

Likely

barba < **bar*(*s*?)*d^h-* : Suggested by **d* in PGm., PSlav., PBalt.

caulae < **kag^h-* : Guaranteed by Lat. *cohum*, **g* in PCelt., PGm.

Plausibly

columba < **ke*/*ol-o*/*umb^h-*: Suggested by PGm. **kulubrōn-*

corbis < **korb^h-*: Suggested by PGm. **krebo-*

In several other cases, the analysis is more complicated. The *b* of Lat. *plumbum* can be reconstructed with a **d^h* like for PCelt. **(ϕ)loudio-*, or it perhaps corresponds to the β of e.g. Gk. μόλιβδος. *Sabulum* could reconstruct to a **b^h* if that is what caused the gemination of Gk. ψάμμος and underlies the *w* of Arm. *awaz*. The analysis of *cucurbita* depends on how the elements align with those of PGm. **hwehwetjō-*. *Sulpur* requires **p* but some Romance forms reconstruct to **b^h*, making its interpretation less than straightforward. And *badius* points to a reconstruction with **d^h* if its suffix is **i̯o*, rather than **iHo* (**iyo*).

More surprising, or at least informative, is the fact that some Latin loanwords with Mediterranean comparanda likewise entered Italic before or during the Proto-Italic treatment of voiced aspirates. Critically, this means that Proto-Italic did not split up until after it entered the Mediterranean zone. The data is as follows:

In Initial Position

Likely

fīcus < **d^{h}īk-*: Guaranteed by PGk. **t(^{h})y/wūko-*, Arm. *t^{c}uz*

faber < **d^{h}ab-*: Guaranteed by Arm. *darbin*

hirundō < **g^{h}ir-*: Guaranteed by Alb. *dallëndyshe*

Plausibly

fascinus < **b^{h}ask-*: Suggested by Gk. βάσκανος if not from /β/ or /ν/[536]

frīgō < **b^{h}rīg-*: Suggested by Gk. φρύγω if not from /f/

funda < **b^{h}end-*: Suggested by Gk. σφενδόνη if not from /f/

hedera < **g^{h}ed^{h}-*: Suggested by Gk. κιθάρα if not from /x/

In Medial Position

Likely

laurus < **lag^{wh}-ro-*: Suggested by Gk. δάφνη and δαύχνα

ālium < e.g. *aG^{h}l̥-i̯o-*: Suggested by Gk. ἄγλῑς, PBerb. **agVlum-*

Plausibly

tabānus < **tab^{h}-*: Suggested by It. *tafano*

As with the non-Mediterranean cases, there are a few forms that resist definitive analysis. *Malva* may have had **g^{w}* depending on its exact relationship with e.g. Gk. μαλάχη. The *d* of *bolunda* may reconstruct to **d^{h}* on comparison with Gk. ὄλυνθος depending on how old it is within Latin. *Faex* and *farciō* have no certain comparanda and therefore do not give any indication of the shape of their source forms.

These words (especially *ālium*, *faber*, *fīcus*, *hirundō*, and *laurus*) would thus represent the earliest loanwords taken up into Italic in the Mediterranean contact zone.

4.3.2.2 Early Contact or Widespread Substrate?

As mentioned above, the large number of substrate lexemes attested with comparanda exclusively in geographically Mediterranean languages contrasts highly with a very

[536] Though see fn. 121.

small number of lexemes that show a more Europe-wide distribution. 7 cases in the dataset exhibit the most widespread distribution, with comparanda in (at least) Greek, Germanic, and Celtic. These include, with the additional language groups in which they are irregularly attested, the following:

Lat. *baculum*	Gk.	Gm.	Celt.	PRom.				
Lat. *baiulus*	?Gk.	Gm.	?Celt.					
Lat. *rādīx*	Gk.	Gm.	Celt.				Alb.	
Lat. *rāpum*	Gk.	Gm.	Celt.		Balt.	Slav.		
Lat. *turdus*	Gk.	Gm.	Celt.		Balt.	Slav.		Arm.
Lat. *tilia*	Gk.	?Gm.	Celt.		??Balt.	??Slav.		Arm.
Lat. *plumbum*	??Gk.	??Gm.	Celt.	(Also ?Basque, ?Berb., ??Kartvelian)				

A widespread distribution, especially amongst IE languages, can have two main explanations. The first possible scenario is that a word was borrowed early, shortly after the dissolution of PIE from a (potentially small) source near the PIE homeland into the newly differentiating IE daughter branches. The second possibility is that a word was borrowed late, upon the arrival of the individual PIE daughter languages in Europe (possibly upon reaching their places of eventually attestation, but it cannot be ruled out that words were borrowed somewhere along the way). This scenario has important implications, because it requires the word(s) to have already had a wide distribution in Europe, perhaps due to them belonging to a widespread language (or more probably language family) spoken by Early European Farmers. Genetic studies have demonstrated that agriculture spread through Europe in the Neolithic via demic diffusion, quite possibly from the same starting point in the Aegean (cf. Hofmanová et al. 2016, Shennan 2018: 107). On the other hand, the time between the arrival of agricultural populations in Europe and the arrival of the IE languages amounts to several millennia. Anthony (2007: 80) has postulated on theoretical factors that Neolithic Europe could have been home to up to twenty to forty distinct language *families*.

In the scenario of early loans from a source close to the PIE homeland, we should expect relatively little variation between languages, and especially not within them. The words would have been borrowed into the proto-daughter languages, perhaps even at intermediary subnodes, and should thus mostly follow known sound laws. This does in fact seem to be the case for the group of words represented by Lat. *taurus*, though curiously not for the Latin word itself (which seems to be a loan post-dating the metathesis of **aurV* > *arṷV*). Nearly all branches (Celtic, Greek, and Balto-Slavic) attest to a pre-form **tauro-* except for Germanic, which preserves **teuro-*. (Albanian could reflect either.) Etruscan also attests *e*-vocalism in *θevru-*. This in combination with the lexeme's deep entrenchment in Semitic (potentially as deep as Afro-Asiatic) suggests a very early loan into just barely differentiated Indo-European.

On the other hand, in the second situation of later *in situ* borrowing, we should expect the full force of differentiation. The words would have been borrowed into the separate IE daughter languages from probably differentiated substrate daughter languages. And in fact, this what we seem to find with the list of widely distributed lexemes. Lat. *plumbum* and its comparanda are widely diverse and for archaeological reasons, as mentioned above, it may represent a relatively late Wanderwort. The comparanda of *rādīx* attest to vocalic alternation between *ā* ~ *a* ~ *i* and what may be the reflex of a schwa. For *rāpum*, the alternations are between **p* ~ **b*^{h} as well as *ā* ~ *a* ~ *ē*/*oi* and *a*-prefixation (in Celtic). The comparanda of *tilia* attest to a complex onset cluster (reflecting both **pt* and **tp*) and potentially complex reflexes of the *a*-prefix phenomenon. *Turdus* and its comparanda attest to several alternations, the most prominent of which include the vacillating presence of a sibilant both at the onset and in the interior of the word and **t* ~ **d* ~ **d*^{h} alternations. *Baculum* and *baiulus* are Italic-Celtic-Greek-Germanic isoglosses whose interpretation is complicated; especially the latter, which seems to have been borrowed multiple times in several branches (Italic: *fascis* ~ *baiulus*, Germanic: **pakk-* ~ **bagg-*, Gk. φάσκωλος ~ φάκελος).

Another indication that Europe may have been home to a large group of related languages is the distribution of the alternating morphological features identified in §3.3. The *a*-prefix especially appears in two lexemes from the list of the most widely distributed (*rāpum* and *tilia*). If the mechanism behind the *a*-prefix is truly as simple as an alternation between a root in the "full-grade" and an *a*-prefixed root in the "zero-grade", then theoretically all cases could have been borrowed at an early date into the just barely differentiated daughter branches; there would be two variants, as with **tauro-* and **teuro-*. However, there are indications that the mechanism was not this simple. The comparanda of *tilia* (especially if ultimately related to *pōpulus*) demonstrate this for instance. It seems more likely that the IE daughter languages borrowed the variants after their dispersion in Europe. Such is also the case for the widely distributed *n*-suffix, especially given that its vacillating presence is not the only irregular alternation occurring in the lexemes where it is found.[537]

Interestingly, there is a chance that both morphological features have a distribution that extends into the Mediterranean. Neither of these is represented in the distinct Mediterranean substrate language identified in §4.2.2.5. This may be due to the sample size, but it seems quite likely that this means there were at least two contact languages in the Mediterranean. For the *a*-prefix, Schrijver (1997: 310) cautiously suggests that cases of (mainly) Greek *prothetisches* α (cf. Furnée 1972: 368-74 with lit.) involving the vacillating presence of *a-* without vocalic reduction of the root could be a related phenomenon. For the *n*-suffix, cases like Lat. *urna*, *orca* ~ Gk. ὕρχη and Lat. *laurus* ~ Gk. δάφνη, δαύχνα are attested with a Mediterranean distribution. For neither of these can it be confirmed that the Mediterranean cases represent the same phenomenon. But it

[537] The extent of the *filix-fulica-sōrex* type velar suffix is complicated by the inability to confirm its relationship with other cases of velar suffixes.

is difficult to rule out, especially for the *n*-suffix. This makes such cases difficult to stratify. If the *a*-prefix and *n*-suffix were found in a Europe-wide distribution including the Mediterranean, both Italic and Greek could theoretically have received their lexemes with these features upon arriving in the Mediterranean.

Thus an important further question remains whether there exist any indications of lexemes that were borrowed along the way to the areas in which the languages would come to be attested, i.e. traces of their prior homelands. Based on the distribution alone, there is indeed a group of words attested in Latin, Germanic, and Celtic to the exclusion of Greek as well as groups including Latin-Celtic and Latin-Germanic isoglosses. Are these due to coincidental lexical loss, or do they represent different contact situations, such as in Figure 4.8?

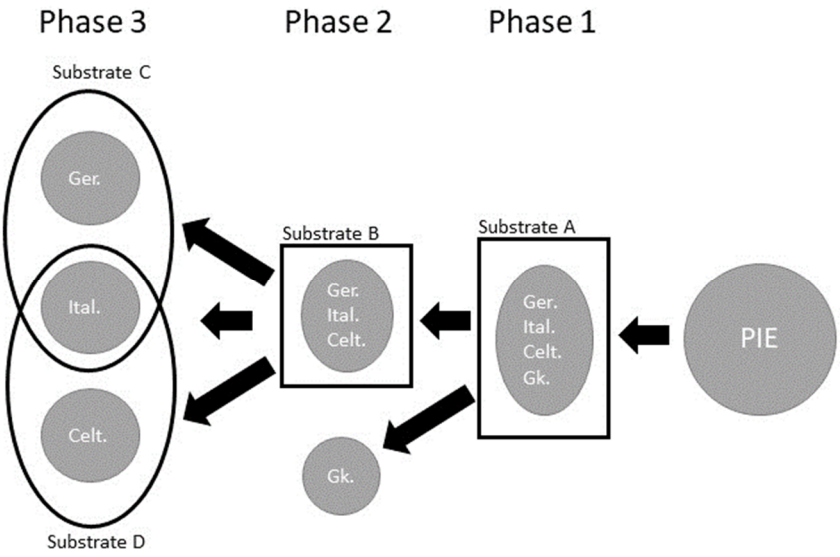

Figure 4.4 Separate contact situations suggested by the distribution of non-inherited lexemes

4.3.2.3 Intermediate Contact Situations

4.3.2.3.1 Stratum Excluding Greek

There is a group of 12 words attested beside Latin in Celtic and Germanic to the exclusion of Greek (*caput*, *catulus*, *caulae*, *corbis*, *far*, *fulica*, *hasta*, *merula*, *nux*, *raudus*, *trabs*, *ulmus*). Of these, *caput*, *corbis*, *fulica*, *hasta*, *merula*, and *nux* are relatively securely Latin-Celtic-Germanic isoglosses. *Merula* attests to *a*-prefixation (for which there are examples of Greek participation as well) and the velar of *fulica* corresponds to Germanic in the same way as other lexemes (*filix* and *sōrex*) that have

Greek comparanda, suggesting that these words were borrowed from a language with which Greek was at one point in contact as well. The alternations attested for *corbis* and *hasta* are mainly of voicing and aspiration, types also attested in lexemes for which there are Greek comparanda. It is *caput* and *nux* that furnish the best evidence of a substrate language with which Greek was never in contact. Their comparanda show a dental ~ velar alternation that is further attested in at least Baltic (van Sluis fthc., see §3.2.1.2.7) but not Greek. The only case that might have a Greek comparandum is that of Lat. *pix* ~ Gk. πίτυς and their comparanda.[538] Given the semantic relationship between pitch and pine trees, the connection is attractive. It would be more certain if both meanings were attested for both forms. As it stands, all forms meaning pitch can be reconstructed to **k* and for pine to **t*. Thus, the lexemes may truly be unrelated leaving Greek with no examples of a dental ~ velar alternation.

This suggests a situation in which Italic, Celtic, and Germanic (along with at least Baltic) were in contact with a substrate language of Europe with which Greek was not in contact. When exactly this would have occurred depends on how early or late the loanwords in §4.3.2.2 entered Greek. Considerations on Italo-Celtic make this a complicated question to answer.

4.3.2.3.2 Italo-Celtic Isoglosses and the Italo-Celtic Subnode

There are three lexemes in the dataset that are attested in Italic and Celtic but not Germanic or Greek. These are *badius*, *sappīnus*, and *bāca*. The first two are Italo-Celtic isoglosses while the latter has only potential Berber comparanda (unless it is ultimately related to Lat. *vaccīnium*). This small number (3% of the data) is probably significant. It may in part be due to lexical loss in the Insular Celtic languages, our best source of Celtic lexical material. But it may well also have to do with the potentially reconstructible Proto-Italo-Celtic subnode. Considerations on the reconstruction of an Proto-Italo-Celtic subnode have important consequences for the stratification of the data. If Proto-Italo-Celtic persisted throughout the period in which loanwords were entering the Indo-European daughter languages, then forms with irregular alternations should be few and due to post-split borrowings. The cases of *badius*, *sappīnus*, and *bāca* are indeed few. The latter two are also plausibly late loans since they attest to irregular alternations even within Italic (though they may have been borrowed at different times, see §4.3.1.2). Some Italo-Celtic isoglosses that reconstruct to the same proto-form (and thus are not included under §2.2.2) may be the result of substrate words borrowed into still unified Italo-Celtic (cf. *caelum*, *hirūdō*; see Stifter fthc. on the latter).

The picture presented by the rest of the Celtic data however seems paradoxical. Of 26 words for which Celtic attests comparanda, only 5 cases (*caput*, *caulae*, *fascis* [as a

[538] This also relies on Gk. κεφαλή 'head' < **g*^{h}*eb*^{h}*h*$_2$*-l-* being only coincidentally similar to the *caput* group. The presence of a cognate in Tocharian (Toch. A *śpāl* 'head', demanding the reconstruction of a laryngeal that the Greek does not otherwise require) is indeed strongly indicative of an unrelated, inherited formation.

relative of *baiulus*], *hasta*, *merula*) can be reconstructed to the same pre-form in Proto-Celtic and Proto-Italic. Most of these are in the Italic-Celtic-Germanic layer excluding Greek which, as argued above, seems likely to be more recent than the layer including Greek. For *tilia* and *rāpum* for instance, Celtic attests to forms with an *a*-prefix against all others. This would imply a contact situation like in Figure 4.9, where Proto-Italo-Celtic and Proto-Germanic had migrated into contact with a language with which Greek did not have contact before reaching their eventual places of attestation and receiving loans from a pan-European substrate language *in situ*. As mentioned above, if the distribution of the *a*-prefix and *n*-suffix includes the Mediterranean, this situation is possible. The Mediterranean loans in Latin would thus have entered Italic in approximately the same region as the other loans, suggesting much linguistic diversity.

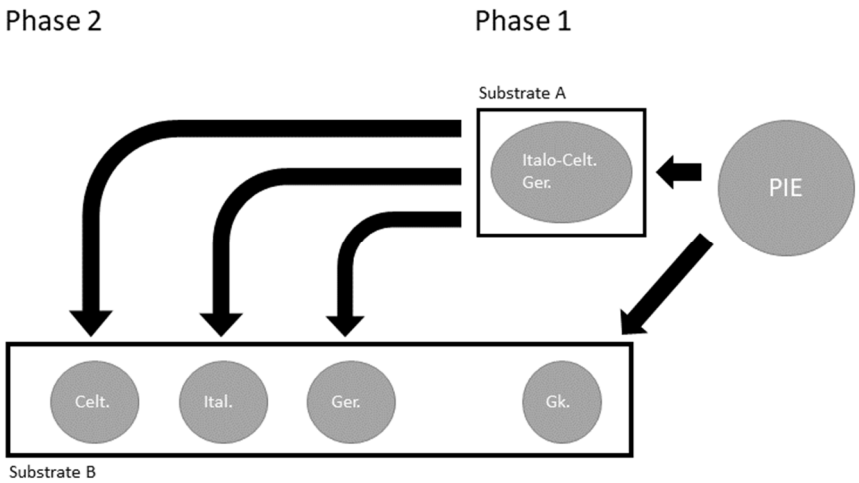

Figure 4.5 One interpretation of contact incorporating Italo-Celtic considerations

Some support of this alternative timeline is given by Stifter (fthc.), who notes that the loss of QPIE **p* in PCelt. **kaφuto-* (belonging to the stratum excluding Greek) makes it look older than PCelt. **arbīno-* (with widespread comparanda including Greek) if the **b* of the latter indeed represents an attempt at rendering a /p/ after that sound was lost in Celtic. Given the difficulty in being able to confirm this (note the warning of Huld 1990: 394-5 mentioned by Stifter fthc. fn. 24), other potential scenarios are possible.

While an Italo-Celtic subnode is not fully accepted by all, Italic does indeed seem to share more innovative features with Celtic than with other branches (Weiss 2022a: 108). It is not certain how early or late Proto-Italo-Celtic would have split into Proto-Italic and Proto-Celtic. Figure 4.10 illustrates a possibility in which Proto-Italo-Celtic split up early

on. The Italic-Celtic-Germanic lexemes would then have been borrowed in a contact situation that occurred *after* Greek had migrated to an area where this language was not present. Proto-Italic and Proto-Celtic speech communities would theoretically have been in close contact near the Alps at a recent date, and the small number of exclusively Italo-Celtic irregular isoglosses would be the result of 1) close proximity to the same

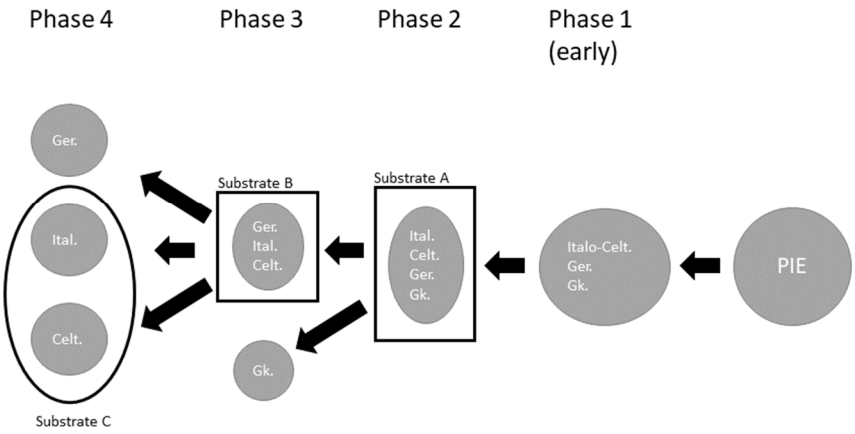

Figure 4.6 An alternative scenario with an early Italo-Celtic split and later close contact

substrate (resulting in many loans entering the languages the same way, mimicking Italo-Celtic unity), 2) lexical loss, and 3) this contact occurring in a part of Europe where there were few speakers of non-IE languages left (cf. Stifter fthc. on the latter point).[539] After subsequently entering the Mediterranean region, Italic re-entered a situation where it was in contact with some of the same languages as Greek.

It remains difficult to decide which scenario best represents the contact situations that resulted in the attested data. But the important results remain: very few cases of non-IE loanwords into Italic and Celtic can be reconstructed to the same proto-form and there are very few cases of exclusive Italo-Celtic irregular isoglosses.

4.3.2.3.3 Italo-Germanic Group

The comparanda shared between Latin and Germanic to the exclusion of Celtic and Greek (*alnus*, perhaps *avēna*, *barba*, *cucurbita*, *excetra*, *faba*, *ferrum*, *fracēs*, *raia*,

[539] Noting cases of alternation within other branches (like Lat. *caulae* ~ Lat. *cohum*; Lat. *trabs* ~ Osc. **tríibúm** ~ U *trebeit*; PGm. **habuda*- ~ PGm. **haubuda*-) another alternative is that Proto-Italo-Celtic could have existed in theory up to a recent date, with irregular alternations between comparanda being the result of borrowings from highly regionally diversified dialects of the contact languages. This is the most problematic possibility, as alternations within branches are generally assumed to represent the latest borrowings.

sulpur) often have a strong connection to differentiated Baltic and Slavic (*alnus*, *avēna*, *barba*, *excetra*, *faba*, *fracēs*). That they demonstrate several types of substrate alternations that are found elsewhere (*n*-suffix: *alnus*, *avēna*, *faba*; *s* insertion: *barba*, *fracēs*; *SK* metathesis: *excetra*) makes their analysis difficult. How many such cases are due to lexical loss in other branches that could originally have attested them? How many are due to contact with separate languages that are nonetheless related to other substrate languages of Europe? *Raia*, *sulpur*, and possibly *cucurbita* are Italo-Germanic isoglosses.

4.3.2.3.4 Germanic Loans in General

It is interesting to note that most of the Germanic comparanda, where they contain a diagnostic phoneme, can be shown to have entered Germanic before the operation of Grimm's and Verner's Laws. It is a pre-Proto-Germanic stage when their consonantism can be reconstructed to match that of the other comparanda most closely:[540]

Lat.	PItal.	QPIE for PGm.	PGm.	Other Comp.
aper	**a**p**ro-*	**h₁e**p**-ŕ-*	**ebura-*	PGk. **e**p**ero-*
ardea	**ar**d**-*	**h₂/₃er**d**-*	**artō(n)-*	PGk. **erō**d**-*
baculum	**b**a**k-*	**b**H**k-́*	**pagjō-*	PGk. **b**a**k-*
caput	**k**a**p**u**t-*	**k**H**p**u**t-*	**habuda-*	PCelt. **k**a**φ**u**to-*
caulae	**k**a**χ-*	**ko**g**ʰ-*	**haga(n)-*	PCelt. **ka**g**yo-*
faba	**b**ʰ**ab-*	**b**ʰ**h₂eu-n-*	**baunō-*	PSlav. **b**ò**bъ*
far	**f**a**rs-*	**b**ʰ**Hr(V)s-*	**bariz-*	PCelt. **b**a**ragi-*
ferrum	**f**e**rsom*	**b**ʰ**rOs-*	**brasa-*	
pannus	**p**a**nno-*	**p**H**n-*	**fanan-*	PGk. **p**ā**no-*
rādīx	**wr**ā**d-*	**u̯reh₂**d**-*	**wrōt-*	PCelt. **wra**d**-*
raudus	**rau**d**o-*	**h₂eru**d**-*	**arut-*	
trabs	**t**r**a**b**-*	**t**r**b-*	**þurpa-*	PCelt. **t**r**e**b**-*
turdus	**t**o**/ur(z)**d**o-*	**t**r**os**d**-*	**þrastu-*	PCelt. **t**r**oz**d**i-*

Table 4.1 Consonantism correspondences between Germanic and other branches

Importantly for the nature of the irregular velar suffix (§3.3.3), the correspondence between Lat. *fulica* ~ OHG *belihha* and Lat. *filix*, *felix* ~ PGm. **brekna(n)-* is best reconstructed to initial QPIE **b**ʰ***, requiring a borrowing before the operation of Grimm's Law. Thus the velar suffix that entered Latin and Greek with **k* entered Germanic as voiced **g*.

In a similar vein, PGm. **hadelō-* and **hadnō-* must have entered Germanic pre-Grimm

[540] A note of caution is that it cannot be ruled out that reconstructions involving the operation of Verner's Law might be misinterpretations. PGm. **pagjō-* can be reconstructed to QPIE **bHk-* to match the Italic and Greek comparanda (as in Table 4.1), but the geminate **kk* of PCelt. **bakko-* corresponds to PGm. **g* < QPIE **g**ʰ*** in e.g. PGm. **managa-* 'many' ~ PCelt. **menekki-* 'abundant'. On the other hand, the correspondence is with PGm. **k* < QPIE **g* in PGm. **balikōn-* ~ PCelt. **bo/ula/okk-* 'coot'.

as Italic and Celtic both attest a reflex of PIE **k*. But as to their second consonant, do they go back to **kHd^{h}-* to match PCelt. **kadVlot-* or to Verner variants of **kHt-* to match PItal. **katVlo-*? In light of PCelt. **arbino-* and Gk. ῥάφυς, PGm. **rōbjōn-* may have entered with **b^{h}* or with **p* (like all other comparanda).

PGm. **kulubrōn-* may attest to QPIE **g* like its Balto-Slavic comparanda, or it may represent a post-Grimm borrowing of **k* like Lat. *columba*. Two cases indeed suggest that a word entered Germanic both before *and* after the operation of Grimm's Law: PGm. **pakka-* (< QPIE **bHG-*) ~ PGm. **bagg-* and PGm. **hrep-* (< QPIE **kreb-*) ~ PGm. **krebō-*.[541] The alternation between PGm. **rugg-* and PGm. **rehhōn-* cannot be explained this way.

That no Germanic comparanda of Latin words securely and exclusively attests to a post-Grimm borrowing suggests that Italic ceased to have contact with the same substrate languages as Germanic before the operation of Grimm's Law. This comes as no surprise if Grimm's Law operated in the mid-first millennium BCE.

4.3.2.3.5 Other Groups

There are 4 words attested in Italic, Celtic, and Greek to the exclusion of Germanic (*caballus*, *calpar*, *lapis*, *paelex*). *Caballus* has widespread attestation that points to a more recent Wanderwort. *Calpar* has a Mediterranean distribution save for its Celtic attestation, and van Sluis (fthc.) proposes it might be a Wanderwort into Celtic transmitted by Etruscan. *Paelex* and *lapis* seem also to have been in currency in the Mediterranean (*paelex* perhaps more so given its potential borrowing into Semitic); their presence in Celtic has unclear implications. The Hallstatt and later La Tène cultures indeed maintained trading networks with Greeks and Etruscans (cf. Kristiansen 1998: 292-3), but this seems too late for the Proto-Celtic loss of **p* in all of these words. Otherwise, there seems to be little evidence of Celtic participation in the Mediterranean substrate(s).

Several other words are attested in Italic, Germanic, and Greek to the exclusion of Celtic (*ascia*, ?*columba*, *ervum*, *filix*, *fungus*, *sabulum*, *sōrex*, *aper*, *ardea*, *pannus*, *viscum*). It is unclear why Celtic does not have these. If the number of exclusive Italo-Celtic irregular isoglosses is so small due to lexical loss (i.e. replacement), then perhaps this is the case here too: the Celtic languages may well have replaced what would otherwise be loans from the same contact situation that produced widespread loans like *rāpum* and *tilia*. This is suggested by at least *filix* and *sōrex* (whose irregular correspondence between Latin and Germanic is shared by *fulica*, for which there is a Celtic comparandum) and *ardea* (with an *a*-prefix, a pattern with examples elsewhere in Celtic). The *SK* metathesis of *ascia* and *viscum* also occurs in *mūlus* and possibly in *tamarix* and *excetra*, also

[541] For the latter form, evidence against a later borrowing is the great variation within the Proto-Germanic forms such that they point to an ablauting *n*-stem, making them archaic (Kroonen 2011: 179-82, 2013: 303).

without Celtic comparanda. But PGm. **þahsu-* ~ PCelt. **tazgo-*, **tasko-* 'badger' makes it difficult to propose that this is the result of a contact situation in which Celtic did not take part.

4.4 A More Inclusive Visualization

The importance of Greek, Celtic, and Germanic for the major distributional tendencies is clear, but this tripartite representation does not effectively represent the placement of all the other languages for which comparanda exist. This is most easily visualized with the assistance of multidimensional scaling.[542] Plotted in the following diagrams are the 61 words from §2.2.2.1 *Non-inherited Origin is Probable*.

The plots are Principal Component Analyses (PCAs) using the lexemes as data points (rows) and their languages of attestation as properties (columns). This takes the form of presence (1) and absence (0) data. However, a traditional PCA treats two identical values as the same; thus two zeros would be interpreted in the same way as two ones. This is clearly inaccurate. (The absence of one particular lexeme in two or more languages does *not* mean that they share an alternative lexeme in its place. In reality, each language that does not have the lexeme in question usually has a different alternative lexeme.) In this way, the dataset is similar to datasets on species abundance in the field of Ecology, where Legendre and Gallagher (2001) recommend several transformations of the data before it is put into a PCA. A Hellinger transformation is suited for data like this with low counts and many zeros. Thus the presence-absence tables were imported to R, Hellinger transformed ("vegan" package decostand("hellinger")), visualized ("ggplot" package autoplot), and re-labelled ("ggrepel" package geom_text_repel). Because the discussion of the lexemes involved uncertain comparanda, the analysis was performed twice: a loose plot (with the uncertain comparanda included) and a strict plot (with only secure comparanda included).

Figure 4.11 displays the loose analysis, in which Principal Components 1 and 2 account for 50.74% of the variation. The vectors (arrows in red) help visualize in space the importance of the major effect of Greek as opposed to Germanic and Celtic. However it also visualizes the effects of Baltic and Slavic.

[542] The statistical portion of this section has been greatly informed by discussions with Paulus van Sluis.

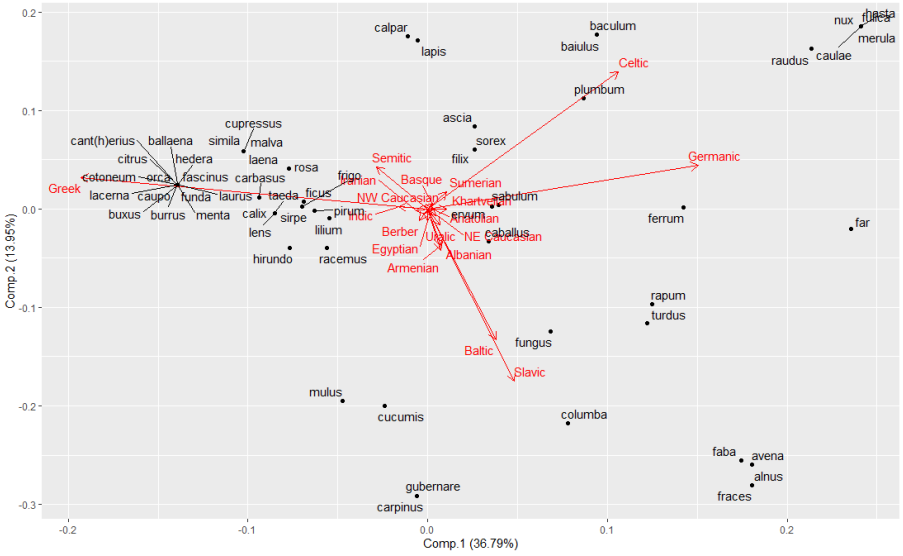

Figure 4.7 PCA of loose analysis

Figure 4.12 contains the strict analysis, in which Principal Components 1 and 2 account for an increased 60.79% of the data. The overall trend remains the same as in the loose plot, with a similar importance played by Greek, Celtic, Germanic, Baltic, and Slavic. Very visible is the Mediterranean cluster, with a major role in placement played by Greek, along with Semitic. The widespread nature of the rest of the data points matches many of the conclusions from above: that there are likely several other contact situations represented by the data. How many different situations are represented as opposed to how many different clusters actually attest to singular large contact situations with subsequent lexical replacement is difficult to confirm.

While the PCAs assist in the visualization of distributional trends, a major caveat is indicated by *plumbum*, *baculum*, and *baiulus* sharing the same position in Figure 4.12: Two lexemes in close proximity share similar distributions between languages, but this is not proof they entered Latin as the result of the same contact situation or that they arose in the same source language.

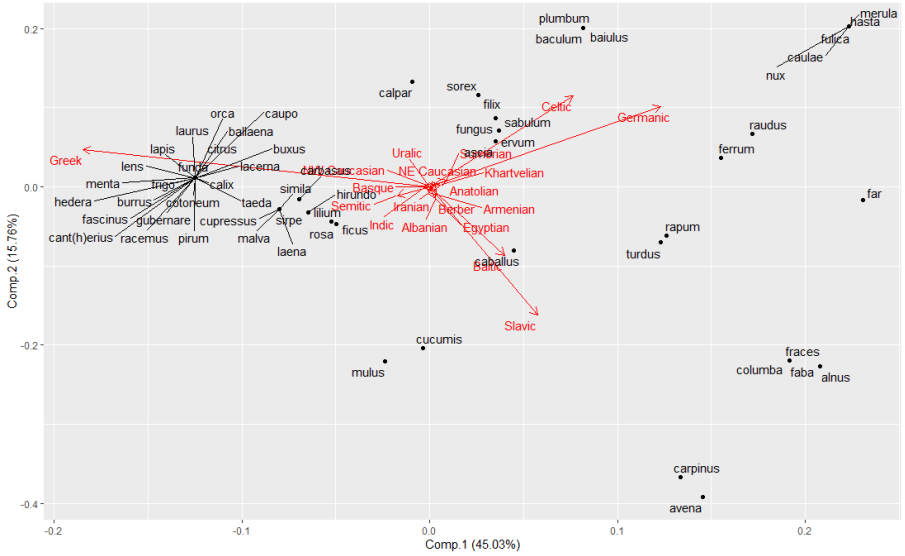

Figure 4.8 PCA of strict analysis

4.5 Summary of Stratigraphy

Good evidence of the very earliest loanwords is partially obscured by the methodology of this thesis, in that loanword status is based on irregular alternations between comparanda. The earliest loanwords would show little variation. *Taurus* is one example, but it happens to be a more recent loan into Latin.

After the split of the Indo-European daughter languages, the stratification of contact situations directly following is difficult and depends in part on whether Proto-Italo-Celtic split very early or late. If early, loanwords with a widespread distribution including Greek are likely some of the earliest. This is followed, as indicated by weak evidence (a velar ~ dental alternation), by a period where Italic, Celtic, and Germanic entered a contact situation with which Greek was not a part. If Proto-Italo-Celtic persisted to a more recent date, then the loanwords with a widespread distribution of attestation may in fact be due to late, *in situ* borrowings from a very widespread, entrenched network of languages. In any case, Italic was also in contact with the same substrate language(s) as pre-Grimm's Law Germanic and separate Baltic and Slavic.

There are very few demonstrable cases of loanwords shared exclusively with Celtic, a fact that must be considered along with the details of the dating of Proto-Italo-Celtic.

Proto-Italic reached the Mediterranean region, at which time it began to borrow lexemes from the same languages as Greek as well as receive loanwords from Greek that were transmitted indirectly through other languages. There is evidence of at least one distinct substrate language of the Mediterranean, along with the possibility of others. Borrowings from these as yet identified sources into Italic languages in the Mediterranean region continued from the period of Proto-Italic until at least the third century BCE. At some point, it seems that some varieties of Latin reborrowed some of the same words at a later date, possibly resulting in Romance forms that attest to geminates.