



Universiteit
Leiden
The Netherlands

Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems

Warries, A.R.

Citation

Warries, A. R. (2022). Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems. *Indo-European Linguistics*, 10, 169-213. doi:10.1163/22125892-bja10022

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3664957>

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

Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems

Abel Radu Warries | ORCID: 0000-0001-6116-6096

Leiden University, Leiden, The Netherlands

a.r.warries@hum.leidenuniv.nl

Abstract

During their migration from the Eastern European steppes to the Tarim Basin, the ancestors of the Tocharians must have come into contact with speakers of different languages, which may have influenced the early Tocharian language. Early Uralic has been identified as possibly having been the source of such influence, especially in the domain of phonology and nominal morphology. In a 2019 article, Michaël Peyrot focused specifically on pre-Proto-Samoyed influence on Tocharian, proposing among other things a comparison of the vowel systems. I will discuss this comparison and give an alternative interpretation. Three difficulties remained with Peyrot's comparison regarding details of 1) the relative chronology of Tocharian sound changes, 2) the mechanism of change, and 3) the relative chronology of sound changes in Samoyed. After addressing these problems in more detail, I conclude that a different vowel comparison is possible, so that the hypothesis that pre-Proto-Tocharians were in contact with pre-Proto-Samoyed substrate is still plausible.

Keywords

Tocharian – Uralic – Samoyed – substrate – vowel system – relative chronology

1 Introduction

It has been proposed by several scholars that the ancestors of the Tocharians were at some point in contact with speakers of early Uralic (e.g., Krause 1951, Schmidt 1990, Bednarczuk 2015, Peyrot 2019). This contact would have

taken place before the Tocharians arrived in the Tarim Basin in north-western China, where their language is attested from the 5th century CE. Since little is known for certain about the whereabouts of the Tocharians before that time, linguistic evidence of contact with another language family like Uralic could provide us with further insights into the prehistory of this branch of Indo-European. If we can establish that early speakers of Tocharian and Uralic encountered one another based on the linguistic data available to us, we should then also be able to determine where approximately this encounter took place, and when. This could shed further light on the migration of the Tocharians from the Indo-European homeland. Arguments advanced in favour of contact mainly involve Tocharian being influenced by Uralic, both in terms of phonology and morphology, and rely on typological comparisons between the two. However, not all of these arguments have been worked out to their fullest extent, and particular care needs to be given to the time depth of the comparison.

Recently, Peyrot (2019) investigated (among other things) the vowel system of Tocharian in light of the Uralic substrate hypothesis, although he did not use Proto-Uralic itself for comparison. Instead, he compared an earlier stage of the vowel system found in the Samoyed branch of Uralic with the pre-Proto-Tocharian vowel system. In the present article I will elaborate on this comparison between the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems, to see if it holds as an argument for contact under closer scrutiny. After discussing some of the remaining problems with Peyrot's proposed comparison (section 2), I will give separate overviews of the relative chronology of phonological changes that occurred in the vowel system of each language group (section 3 for Samoyed and section 4 for Tocharian). A thorough understanding of the relative chronology is necessary to avoid anachronisms in the eventual comparison. Based on these relative chronologies I will argue that a comparison is still possible; in fact, two slightly different comparisons present themselves (section 5). In the discussion (section 6) I will determine which comparison holds up better, considering the currently available evidence, and how they relate to hypotheses from archaeology and genetics about the prehistory of the Tocharians. In the rest of this introductory section, I will give some further background to the problem at hand.

According to the current consensus, the Indo-European homeland was in the Ponto-Caspian steppes of Eastern Europe, far to the west of the region where Tocharian is attested. Thus, a natural question is how the speakers of Tocharian made their long journey eastward. Currently the best scenario on archaeological grounds is based on the idea that the Afanasievo Culture north and

north-west of the Altai and Sayan mountains represents the ancestors of the Tocharians. Afanasievo shows a close match with the Yamnaya steppe culture associated with Proto-Indo-European, and might have been an intermediate stopping point for the ancestors of the Tocharians before they ended up migrating further to the south and into the Tarim Basin. The linguistic support for this connection between the Afanasievo Culture and Tocharian has not been fully worked out, however (see, e.g., Anthony 2007, Anthony 2013, Mallory 2015, Kroonen, Barjamović, & Peyrot 2018).

As mentioned above, Tocharian has already been investigated for traces of influence from Uralic. In Tocharian morphology, the rise of an agglutinative nominal case system has been connected with the likewise agglutinative Uralic languages by a number of scholars (e.g., Krause 1951, Schmidt 1990, Bednarczuk 2015, Peyrot 2019). A more striking Tocharian change is found in the phonology, with the disappearance of all voicing distinctions that were present in the Proto-Indo-European stop system (e.g., PIE $^{*}/k̥\ g̊\ gʰ/ > PT\ ^{*}/k/$). Since Proto-Uralic is also reconstructed without a voicing distinction in its stop system, it has been proposed that the Tocharian stop system was influenced by Uralic (Ivanov 1985, Kallio 2001: 224–226, Bednarczuk 2015, Peyrot 2019).

To illustrate the similarities, the developments of the stops from Proto-Indo-European to pre-Proto-Tocharian are given in Table 1, next to the typologically similar Uralic system. This table does not include the palatalisation opposition that eventually arose in Tocharian, as that is not directly paralleled in the earliest forms of Uralic. In light of this comparison, it may be noteworthy that the sequences PIE $^{*}t_{\check{i}}$ and $^{*}d^{h}_{\check{i}}$, as well as PIE $^{*}d$, normally yielded PT $^{*}t^s$. When comparing the Tocharian stop system to the Uralic stop system, one could posit that PT $^{*}t^s$ developed from an earlier palatal or postalveolar affricate that could be seen as similar to PU $^{*}\acute{c}$ ($^{*}[t_{\check{c}}]$) or $^{*}\acute{\check{c}}$ ($^{*}[t_{\check{s}}]$). Especially for the PIE $^{*}t_{\check{i}}$ and $^{*}d^{h}_{\check{i}}$ that yielded this Tocharian phoneme, such an affricate seems like a possible intermediary stage. Changes of several types of sibilant affricates to $^{*}t^s$ are quite common (see Kümmel 2007: 198–202 for examples), and within Uralic a change from PU $^{*}\acute{\check{c}}$ to t^s is known from Saami.¹

The consonant systems of Tocharian and Proto-Uralic are clearly very similar, but the vowel correspondences are much less straightforward. The rest of

1 At some point in the prehistory of Tocharian, front vowels palatalised preceding consonants, so that in attested Tocharian there existed a consonant c $[t_{\check{c}}]$. This is a separate phoneme from the PT $^{*}t^s$ under discussion here. Positing something like a pre-PT $^{*}\acute{\check{c}}$ or $^{*}\acute{c}$ as a pre-stage to PT $^{*}t^s$ does not come into conflict with the Tocharian c , since a secondary change to $^{*}t^s$, as hypothesised here, can easily be posited before palatalisation took place.

TABLE 1 Typological comparison between the
pre-Proto-Tocharian and Proto-Uralic stops

PIE	(pre-)PT	PU
p, b ^h	> p	p
t, d ^h	> t	t
tj, dhj, d	(> č/ć?) > t ^s	č, ć
k, g, g ^h	> k	k
k ^w , g ^w , g ^{wh}	> k ^w , k [*]	— ^{**}

* Depending on the environment (see, e.g., Kim 1999).
** Pystynen (2017) has suggested that *k^w perhaps existed
at some point in pre-Proto-Samoyed (cf. also 3.3.)

this article will be devoted to a detailed study of the available evidence for the reconstruction of the vowel systems of both pre-Proto-Tocharian and pre-Proto-Samoyed, and a discussion of the resulting comparison of these vowel systems.

2 Difficulties with the previous comparison

In this section I will summarise the vowel comparison as it was advanced by Peyrot, and comment on some remaining difficulties to be addressed in the present article. Peyrot (2019: 81–89) reconstructs a pre-Proto-Tocharian vowel system with seven vowels */i e a o u ɐ̯² and compared this with a pre-Proto-Samoyed vowel system with the vowels */i e a o u ɛ j/.³ These systems represent an intermediate reconstruction between Proto-Indo-European and Proto-Tocharian on the one hand, and between Proto-Uralic and Proto-Samoyed on the other. This is illustrated for both languages in Table 2, in which pre-Proto-Tocharian is given between post-Proto-Indo-European and Proto-Tocharian, and pre-Proto-Samoyed between Proto-Uralic and Proto-Samoyed, both as reconstructed by Peyrot.

Based on the evidence of the stop system and the case system, combined with the similarities between the pre-Proto-Tocharian and the pre-Proto-Samoyed vowel systems thus reconstructed with the same configuration of

2 This notation differs from the one used by other scholars: *e* = *æ*, *a* = *ā*, *o* = *ǫ*, *ɐ̯* = *ä*.
3 In Uralic notation, *ɛ* represents a mid, unrounded back vowel, and *j* a high, unrounded back vowel.

TABLE 2 The development of the Proto-Tocharian and Proto-Samoyed vowel systems according to Peyrot (2019), each with an intermediary stage with seven vowel phonemes in the middle column

Post-PIE			Pre-PT			PT		
ei ī i		u eu	i	ə	u	i	ə	u
ē e		o ō	e	ĕ	o	e		o
	a ā			a			a	

PU			Pre-PSam.			PSam.		
i ü		u	i	ĭ	u	i ü	ĭ	u
e	ɛ ^a	o	e	ɛ	o	e ö	ɛ	o
ä	a			a		ä		â
						reduced: â		

a This Proto-Uralic vowel can also be written as **ĭ*; I will use the notation **ɛ* throughout this article. Regardless of the notation adopted for Proto-Uralic, PU **ɛ/*ĭ* was a single phoneme, which split in Samoyed to yield PSam. **ɛ* and **ĭ* as distinct phonemes.

seven vowels, Peyrot concludes that pre-Proto-Tocharian was likely at some point in contact with, and influenced by, pre-Proto-Samoyed. However, there are three main difficulties with Peyrot’s vowel comparison, mostly already remarked upon by Peyrot himself in some form:

1. The pre-Proto-Tocharian vowel system used in the comparison is anachronistic, in particular with regards to palatalisation (cf. Peyrot 2019: 91) and umlaut involving PIE **u*.
2. The mechanism by which the (post-)Proto-Indo-European vowel system would be changed into pre-Proto-Tocharian due to a pre-Proto-Samoyed substrate remains unclear, and extra independent developments would be required to “prepare” pre-Proto-Tocharian for the development of a vowel system like that of pre-Proto-Samoyed (cf. Peyrot 2019: 104–105).
3. The pre-Proto-Samoyed vowel system with seven vowels used for the comparison with pre-Proto-Tocharian by Peyrot (cf. Peyrot 2019: 90) probably never existed.

Sub 1: In his conclusion on the vowel system, Peyrot (2019: 91) writes that the merger of PIE **i*, **e*, and **u* into pre-PT **ə* needed to arrive at the intermediary system with seven vowels also necessitates phonemic palatalisation at

that stage of pre-Proto-Tocharian. This is because in Tocharian the only difference between original PIE **i* and **e* on the one hand and **u* on the other is the palatalisation caused by the former two. At the same time PIE **o* and **ē*, which in Peyrot's reconstruction ended up as pre-PT **ē* and **e* respectively, would not yet have merged as **e*, despite the fact that those two vowels also end up as a neutral vs. palatalising pair in Proto-Tocharian. PIE **eu* would furthermore become a palatalising pre-PT **u*, corresponding to pre-PSam. **u*, so that palatalisation really should have been firmly established at the intermediary stage that Peyrot reconstructs as pre-Proto-Tocharian. A continued phonemic difference between **ē* and **e* may therefore be anachronistic: one would expect that when palatalisation became phonemic, the contrast between **/Cē/* and **/Ce/* became re-phonologised as **/Ce/* vs. **/C^he/* as well. It is more economical to posit a single palatalisation rule, so long as there is no compelling evidence to suppose that there were two.

Peyrot (ibid.) also remarks that the consonant system associated with his pre-Proto-Tocharian is not as good a match for the pre-Proto-Samoyed system, due to the necessity for palatalisation to be already established. This is because pre-Proto-Samoyed had no such pervasive palatalisation opposition.⁴ On account of the comparison of the consonant systems, therefore, it would be preferable to posit palatalisation as a later change in pre-Proto-Tocharian, after any contact with early Samoyed. However, that would not allow for the reconstruction of the vowel system adopted by Peyrot in his comparison with pre-Proto-Samoyed, since the merger of PIE **i*, **e*, and **u* as pre-PT **ə* necessarily resulted in phonemic palatalisation.

An additional anachronism with this merger of PIE **i*, **e*, and **u* in Peyrot's pre-Proto-Tocharian is that PIE **u* caused umlaut of pre-PT **e* (from both PIE **ē* and **o*) to yield PT **o* > TAB *o*, and was itself also affected by a following **o* < PIE **ō* to become PT **o* (Pinault 2008: 431–433, Hackstein 2017: 1321). This means that the umlaut in, e.g., TAB *or* < PT **orə* < **eru* 'wood' should either precede the vowel reduction assumed to have taken place before the rise of Peyrot's pre-Proto-Tocharian vowel system, or be simultaneous with it. In any case, a new PT **o* would be phonologised already at that stage, yielding a system with eight vowels instead of seven, with an extra back rounded vowel. Peyrot dated this change later than the merger of PIE **u* with PIE **i* and **e* as PT **ə* (Peyrot 2019: 83), but that is impossible due to the fact that short PIE **u* still caused *u*-umlaut.

4 Proto-Samoyed had a palatalisation opposition in the nasals: neutral **n* vs. palatal **n̥*. Perhaps also **l* vs. **l̥* existed in earlier pre-Proto-Samoyed, but the Tocharian system is quite different: practically every consonant had a palatal(ised) counterpart.

Sub 2: Proto-Uralic and pre- and Proto-Samoyed had the vowels **i*, **e*, and **u*, just as the earliest pre-Proto-Tocharian would have had inherited from Proto-Indo-European. Peyrot does not compare pre-Proto-Tocharian **/i e u/* with pre-Proto-Samoyed **/i e u/*, however. Instead, he regards the PT **ə* that these three vowels merged into in Tocharian as a counterpart to the pre-Proto-Samoyed high unrounded vowel **i̥*. Peyrot observes that it is unclear how exactly PIE **/i e u/* should have become **ə* ($\pm^*[i̥]$) in Tocharian due to pre-Proto-Samoyed influence, instead of just remaining as **/i e u/*. That is to say, the mechanism by which (part of) the Indo-European vowel system was changed to be like pre-Proto-Samoyed is unclear. This observation prompted Peyrot to conclude that “we have to assume that most of the drastic changes in the vowels had already started off before influence from Pre-Proto-Samoyed took place, and that these were then under the influence of Samoyedic fixed in the form [of the pre-Proto-Tocharian system]” (Peyrot 2019: 104–105). While something like that could have happened, it largely robs the comparison with Samoyed of its explanatory power. In the consonant system the striking loss of voicing distinctions can be better understood when viewed as the result of Uralic or early Samoyed influence, but if the vowel system had already undergone its most significant changes before contact, a pre-Proto-Samoyed substrate scenario does not contribute as much to our understanding of these Tocharian changes.

Sub 3: Peyrot (2019: 88) arrived at a pre-Proto-Samoyed system **/i e a o u ɛ ɨ/* by subtracting the vowels **/ü ö ä ə~ə̃/* from the traditional Proto-Samoyed vowel system (and by writing PSam. **ă* as **a* as in Proto-Uralic). For the most part it is true that the Proto-Samoyed vowels removed by Peyrot to reconstruct pre-Proto-Samoyed are innovative. However, that does not necessarily mean that in the relative chronology of Samoyed an intermediary system existed with seven vowels as Peyrot reconstructs it. It rests on the assumption that a number of vowel changes occurred in pre-Proto-Samoyed before other new vowels were innovated; namely the PU **ɛ* split into **i̥* and **e̥*, as well as the changes of PU **ü* to **i̥*, and PU **e* and **ă* to **i̥* and **e̥* respectively. This has so far not been shown by a detailed study of the Samoyed data.

Additionally, the pre-Proto-Samoyed and Proto-Samoyed vowel systems given by Peyrot are based on the traditional view that PU **e* merged with **i̥* while PU **ă* shifted to **e̥*, which has been disputed in more recent scholarship on Samoyed. Specifically, Helimski (2005) investigated instances where supposed PSam. **i̥* corresponded to Nganasan **i̥* instead of to expected Nganasan **i̥*. His conclusions have important ramifications for the development of the Samoyed vowel system, and thus also for the comparison with Tocharian. The majority of examples of Ngan. *i̥* occur next to a labial, so that was supposed to

be a conditioning factor (Mikola 2004: 76–77), although a few remaining cases had not been properly explained. Helimski (2005) observed that this vowel is written as *e* or *ê* in Old Nganasan material collected by Castrén, and that etymologically it often corresponds to PU **e*. This means that PU **e* was not changed to **i* until after Proto-Samoyed, contrary to what had been previously assumed. A change from **e* to **i* occurred in all individual Samoyed languages except Nganasan, where PU **e* instead became *ɨ* relatively recently. Further obscuring this correspondence is a later Nganasan assimilation of *ɨ* to *i* in palatal contexts (ibid.). On the basis of the Nganasan evidence, the reconstruction of the Proto-Samoyed vowel system has been changed to **/i ü e ö ä a ɨ ɛ ə ɛ̃/* in more recent scholarship. The former PSam. **ä* is now rewritten as **a*, and PSam. **ä* and **e* continue PU **ä* and **e* directly (see Helimski 2005, also Aikio 2006). Thus, one of the changes on which Peyrot's pre-Proto-Samoyed system with seven vowels is based appears not to have occurred, and a system with three (non-rounded) front vowels **/i e ä/* needs to be taken into account for pre-Proto-Samoyed at any time depth.

Peyrot (2019: 88–89) discussed this revision to the Proto-Samoyed system, but remained skeptical. The exceptions to a fully regular change PU **e* > Ngan. **i* seem to be PU **mexi-* 'give, sell' > Ngan. **mi-* (not ***mɨ-*) and PU **sew(i)mä* > Ngan. *timi* 'tooth' (Peyrot 2019: 89 with fn. 16, see also Aikio 2002: 34–35 on this latter etymology). There is also the tentative etymology PU **wexi-* 'take' > PSam. *?*i-* 'id.' (Aikio 2013: 171). All of these examples show a sequence **eCi*, where the consonant **-C-* is a weak one that disappeared in Samoyed. It may thus be supposed that such sequences did develop to simple PSam. **i*. Although the exact manner in which this change took place is unclear to me, it seems that the exceptions to Helimski's re-interpretation of the Proto-Samoyed vowel system based on Nganasan form a coherent group. I therefore accept the updated Proto-Samoyed vowel reconstruction with retention of PU **e* > PSam. **e* as the regular development.

With these changes in our understanding of the relation between the Proto-Uralic and Proto-Samoyed vowel systems, Peyrot's comparison with pre-Proto-Tocharian is less exact. In the next subsection I will go through the development of the pre-Proto-Samoyed vowel system and what is known about its relative chronology in more detail, so as to arrive at what I think are the possible stages of development that can be used for comparison with pre-Proto-Tocharian. The other two difficulties with Peyrot's comparison may then receive possible solutions as well.

3 Development of the pre-Proto-Samoyed vowel system

In this section I will discuss the relative chronology of changes that took place in pre-Proto-Samoyed. A solid understanding of the different stages of development of the vowel system is necessary before making a typological comparison with Tocharian: we need to know what different stages of pre-Proto-Samoyed looked like before we can compare it to pre-Proto-Tocharian. The sound developments that changed the structure of the vowel system are:

- the split of PU **e* > PSam. **i* and **e*;
- the development of PSam. **a* and **â* from PU **i* and **u*;
- the unrounding of PU **ü* > PSam. **i* and the rise of new PSam. **ü*;
- the development of PSam. **a* (formerly written as **ä*).

I will not treat consonant changes in any detail here, as they are mostly irrelevant for the development of the vowel system. The change of many PU **o* to PSam. **â* did not change the number of vowels present in the pre-Proto-Samoyed vowel system, so I will largely leave that aside as well. I will present all the data that I am aware of, so that it is clear on what I base myself.

3.1 PU **e* > PSam. **i* and **e*

The Proto-Uralic back unrounded vowel **e* underwent a phonemic split into two different Proto-Samoyed reflexes: **e* and **i*. These two vowels had become fully differentiated by the time of Proto-Samoyed and are not in complementary distribution. Janhunen (1981: 233–234) and Sammallahti (1988: 484) take the syllable structure as the differentiating factor between the words showing PSam. **i* and those with **e*. The high vowel **i* would be the reflex in open syllables, while **e* occurred in closed syllables. This distribution accounts for such examples as PU **ēla* ‘space underneath’ > PSam. **ilâ* vs. PU **ēpti* ‘hair of the head’ > PSam. **eptâ*. Under this hypothesis, the distribution could only have originated after certain Proto-Uralic clusters were simplified in Samoyed, as in, e.g., PU **meksa* ‘liver’ > **metâ* > PSam. **mjtâ*. On the other hand, geminates should still have been preserved at the time, on account of the Proto-Samoyed mid vowel **e* in PU **kečči-* ‘smelly, rotten, mouldy’ > PSam. **kečâ-* (on this etymology, see Aikio 2014a: 5–8).

Janhunen and Sammallahti ascribed the Proto-Samoyed mid-vowel reflex in certain other words to a vowel sequence derived from a cluster with the “Uralic laryngeal” **x* of unknown phonetic quality. This **x* was originally reconstructed in words where Finnic shows a long vowel: PU **dēxmi* > PSam. **jēâm*; cf. the long vowel in PFi. **tōme-* ‘bird cherry’. Aikio (2012) has shown that it is not warranted to reconstruct a vowel sequence here for Proto-Samoyed (the attested reflexes are SlkTaz *čēm* and Kam. *lem*, both of which can be derived

from a form with a simple vowel PSam. **jem*), and that PU **x* does not need to be reconstructed in this environment for Proto-Uralic. Pre-consonantal PU **x* was reconstructed by Janhunen precisely in order to account for the correspondence between a Samoyed vowel sequence and a Finnic long vowel (e.g., PFi. *kõle* : PSam. **kââ* ‘die’; see Aikio 2012: 231). However, since the Finnic long vowels have been explained differently by Aikio, the Proto-Uralic reconstruction with pre-consonantal PU **x* is obsolete according to the current understanding of Uralic sound correspondences. With the new reconstruction of PU **dʲemi* > PSam. **jem* this word follows the rule whereby **ɛ* results in closed syllables, provided that this rule came into effect after the loss of final **-i*. The nominative **jem* was then a closed syllable, while for instance the accusative **jemâm* would likely still have had an open syllable, so that paradigmatic alternations of the type **jem* : **jemâm* expected to result from the conditioning based on syllable structure should have been analogically levelled in favour of the nominative form.

However, there are a few exceptions that cannot easily be explained by this sound law. PU **ləmpi* ‘pond, swamp’ > PSam. **ləmpâ*, PU **lənti* ‘lowland’ > PSam. **ləntâ* and PU **jɛŋsi* ‘bow’ > PSam. **(j)intâ* all show the high-vowel reflex **i* in a closed syllable. They all have a cluster with a nasal, so that an additional rule can be posited whereby a nasal coda caused the vowel to become **i* as well. An exception to this is PU **peŋka* ‘psychedelic mushroom’ > PSam. **peŋkâ* ‘get drunk’, which retained the mid vowel despite containing a nasal cluster. Additionally, PU **ɛciw-* ‘camp’ > PSam. **ɛso-* should have become PSam. ***iso-* instead, due to its **ɛ* standing in an open syllable. This formation is also reflected in Finnish *asua* ‘live, dwell’, which means that it is likely of Proto-Uralic age (Aikio 2021: 16–17). Perhaps the expected outcome of **ɛdi* ‘year / autumn’ > PSam. **erö* should similarly have been ***irö*, but it is possible that this word was **er* at some point after apocope, before a suffix **-ö* was added. Taking these words into account, Zhivlov states that PSam. **i* obtains in a Proto-Samoyed open syllable and before the clusters **mp* and **nt*, while **ɛ* is found in a Proto-Samoyed closed syllable and in roots only consisting of **CV*. When a vocalic suffix is added to an open stem, **ɛ* is preserved (Zhivlov 2014: 125). Thus, the development of PU **ɛciw-* > PSam. **ɛso-* would be exceptional because there was such a suffix.

Aikio works with a different rule, whereby PSam. **i* is the reflex before a second-syllable PU **-a*. This condition mostly holds up if it occurred before some instances of PU **-a* became **-â* (as in, e.g., PU **meksa* ‘liver’ > PSam. **mĭtâ*). This change to **i* would also occur before a nasal cluster, even in *i*-stems (Aikio 2015b: 34). The only problem is PU **peŋka* ‘psychedelic mushroom’ > PSam. **peŋkâ* ‘get drunk’, which again behaves contrary to expecta-

tion. On the other hand, the development of PU **ećiw-* ‘camp’ > PSam. **eso-* ‘id.’ is fine without the need to posit an exceptional development in suffixed forms. This means that on the one hand PSam. **peŋkâ-* ‘get drunk’ and on the other **eso-* ‘camp’ bear most of the weight of “proving” the respective general conditioning factors—either the open vs. closed syllable or the quality of the second-syllable vowel.

Below I give the examples that I am aware of, starting with those Uralic words that ended up with an **ĭ* in Proto-Samoyed.⁵

- PU **ela* ‘space below’ > PSam. **ilâ* ‘id.’ (Sammallahti 1988: 536)
- PU *±*ena-eppi* ‘parents-in-law’ > PSam. **inâpâ* ‘father-in-law’ (Sammallahti 1988: 536)
- PU **eŋa* ‘tame’ > PSam. **jŋâ* ‘id.’ (Sammallahti 1988: 536; Aikio 2015b: 59)
- PU **ekta-* ‘hang (tr.)’ > PSam. **itâ-* ‘id.’ (Sammallahti 1988: 536)
- PU **meŋsa* ‘liver’ > PSam. **mitâ* ‘id.’ (Sammallahti 1988: 538)
- PU **seksa* ‘Siberian pine’ > PSam. (der.) **tîtâjŋ* ‘id.’ (Sammallahti 1988: 540; Aikio 2015b: 60)
- PU **jeŋsi* ‘bow’ > PSam. **jintâ* ‘id.’ (Sammallahti 1988: 537)
- PU **lempi* ‘pond, swamp’ > PSam. **lîmpâ* ‘id.’ (Aikio 2014b: 86)
- PU **lenti* ‘lowland’ > PSam. **lîntâ* ‘id.’ (Aikio 2014b: 86)

The following words have the mid-vowel reflex **e* in Proto-Samoyed.

- PU **edi* ‘year’ > PSam. **erö* ‘autumn’ (**-ö* is a suffix; Aikio 2012: 233–234)
- PU **ećiw-* ‘camp’ > PSam. **eso-* ‘id.’ (Aikio 2012: 241)
- PU **demi* ‘bird cherry’ > PSam. **jem* ‘id.’ (Sammallahti 1988: 536–537; Aikio 2012)
- PU **jexi-* ‘to drink’ > PSam. der. **er-* ‘id.’, **ekâl-* ‘gulp down’ (Aikio 2006: 38–40)
- PU **ledi-* ‘frighten’ > PSam. **l/nę(â)-* ‘id.’ (Aikio 2014b: 85–86)
- PU **ŋeli* ‘arrow’ > PSam. **ŋej* ‘id.’ (Aikio 2012)
- PU **ŋeri* ‘wetness, dampness; wet place, bog’ > PSam. **ŋer* ‘wet and sticky substance’ (Aikio 2006: 20–21)
- PU **peŋi-* ‘put’ > PSam. **peŋ-* ‘id.’ (Sammallahti 1988: 539)
- PU **epti* ‘hair of the head’ > PSam. **eptâ* ‘id.’ (Sammallahti 1988: 536)
- PU **keččī* ‘spoiled; bad-smelling’ > PSam. **kečâ-* (Aikio 2014a: 5–8)
- PU **lepci* ‘cradle’ > PSam. **l/jepsâ* (UEW: 230)

5 PU **metka* ‘way, detour (?)’ > PSam. **mitâ* ‘way’ (Aikio 2015a: 13–15) would also belong to this category, but Mikhail Zhivlov (p.c., 24.8.2021) has kindly informed me that the Taz Selkup word *miti* ‘way, track’, which constituted the Samoyed side of the comparison, can instead be connected with PSlk. **uâtta* ‘way, track’, and that the connection with Fi. *matka* ‘journey’ (etc.) has now been abandoned also by Aikio.

- PU **peŋka* ‘psychedelic mushroom’ > PSam. **peŋkã* ‘get drunk’ (UEW: 355–356; Zhivlov 2014: 133)

Neither group is completely unified by a single factor. If labial and coronal nasal clusters as in **lintã*, **lĩmpã*, **(j)ĩntã* (after assimilation of the original cluster PU **-ŋs-*) raised the vowel, but the dorsal nasal cluster in **peŋkã* did not, the rule giving **ẽ* in closed syllables holds with the exception of PU **ẽciw-* ‘camp’ > PSam. **ẽso-*. It seems somewhat unattractive to posit different developments caused by **-mC-* and **-nC-* on the one hand and **-ŋC-* on the other, but the possibility cannot be dismissed outright. With the conditioning that **-a* in the second syllable and nasal clusters caused **ĩ* only PSam. **peŋkã* remains exceptional and therefore problematic, while PU **ẽciw-* > PSam. **ẽso-* is regular without requiring a separate rule for suffixed forms.

Since the reflex **ẽ* in PSam. **peŋkã* ‘get drunk’ does not quite behave as expected in both accounts, one might consider the possibility that this word is not directly inherited from PU **peŋka* ‘psychedelic mushroom’. A cognate is also found in Ob-Ugric, with PMs. **pĩŋk* and PKh. **pãŋk* ‘fly agaric (mushroom)’. Based on the tables given by Zhivlov (2007: 284), the correspondence PMs. **ĩ*: PKh. **ã* goes back to Proto-Ob-Ugric **ẽ* in an *a*-stem, meaning that the preform would have been POUg. **pẽŋka* at the Proto-Ob-Ugric stage. This form could conceivably have been adopted into early Samoyed as **peŋkã*, after PU **ẽ* had split into PSam. **ĩ* and **ẽ*. Janhunen (1998: 477) mentions two more possible pre-Ugric borrowings into Proto-Samoyed, *num* ‘sky, god’ and *kãlma* (**kalmã* in Janhunen’s notation) ‘dead body, grave’. Zhivlov (2014: 133) argues against borrowing, as the Ob-Ugric verb of the same meaning as PSam. **peŋkã* ‘get drunk’ contains a derivational suffix: PMs. **pĩŋkal-*, PKh. **pãŋkal-* ‘get drunk’. According to Zhivlov, “there is no reason to suppose either the loss of the suffix on Samoyed soil, or the borrowing of the Ob-Ugric noun as a Samoyed verb” (ibid.). While I agree that those two possibilities are certainly not very likely, it is still conceivable that the noun was borrowed as such into early Samoyed, from which the attested verb was then derived independently. This option is also put forward by Holopainen (2019: 187). On account of the suffix uniquely found in the Ob-Ugric verb, a separate semantic derivation of the type ‘psychedelic mushroom’ → ‘get drunk’ would be needed in any case, even if Samoyed inherited the noun directly from Proto-Uralic, so that both accounts are equally economical in that regard.

3.2 PU **i*, **u* > PSam. **ə*, **õ*

The Proto-Uralic high vowels **i* and **u* are also of interest to us, as they underwent a characteristic development in Samoyed, leading to the emergence of new vowel phonemes. Both PU **i* and **u* have a dual reflex in Proto-Samoyed,

as they are either preserved as PSam. **i* and **u*, or reduced to PSam. **ə* and **â* respectively. There may not have been a phonetic difference between the front **ə* and back **â* anymore in Proto-Samoyed, but as shown by Helimski (1993) for Nganasan and by Salminen (2012) for Nenets, they continued to behave differently with regards to vowel harmony. The conditioning of the reduction from PU **u* and **i* has been described in various ways. Janhunen (1981: 223) gives as a conditioning for PU **u* > PSam. *â* the presence of a PU **-a* in the next syllable. In other words, a sequence of PU **u-a* changed to PSam. **â-â*. This rule captures most instances of the change from PU **u* to PSam. **â*. Sammallahti (1988: 484), essentially following Janhunen, defines the environment in which PU **u* turned into PSam. **â* as before a low vowel in the next syllable (e.g., PU **kupsa-* ‘extinguish’ > PSam. **kâptâ*), or when a second-syllable high vowel was lost and the intermediary consonant was neither a lateral nor **x* (e.g., PU **suŋi* ‘summer’ > PSam. **tâŋ*). The conditioning of this change by second-syllable **-a* is supported fairly well by the data, although there are a few exceptions.

- PU **kura-* ‘bend; crooked, bent’ > PSam. **kârâ-* ‘id.’ (Aikio 2006: 14–15)
- PU **suwd’a* ‘finger’ > PSam. **tâjâ* (Sammallahti 1988: 540)
- PU **juka* ‘small river’ > PSam. **jâkâ* (Sammallahti 1988: 537)
- PU **lupsa* ‘moisture’ > PSam. **jâptâ* (Sammallahti 1988: 538)
- PU **kupsa-* ‘extinguish’ > PSam. **kâptâ-* (Sammallahti 1988: 537)
- PU **mud’a* ‘land, earth, soil’ > PSam. **mâjâ* (Aikio 2002: 22–23)
- PU **muka* ‘back’ > PSam. **mâkâ* (Sammallahti 1988: 538)
- PU **mura* ‘berry’ > PSam. **mârâ* (Sammallahti 1988: 538)
- PU **pud’a-* ‘to hit, split, break?’ > PSam. **pâjâ-* ‘to chop wood’ (Aikio 2006: 22–23)
- PU **muna* ‘egg’ > PSam. **mânâ* (Sammallahti 1988: 538)
- PU **tunta-* ‘teach, accustom to’ > PSam. **tântâ-* ‘id.’ (Aikio 2002: 44–45)
- PU **jupta-* ‘say, speak’ > PSam. **jâptâ-* (Helimski *apud* Aikio 2002: 48–49)
- PU **kuma-* ‘fall over’ > PSam. **kâmâ-* (Sammallahti 1988: 537)
- PU **kuña-* ‘close eyes’ > PSam. **kânâ-* (Sammallahti 1988: 537)
- PU **kura* ‘knife’ > PSam. (der.?) **kârû* ‘id.’ (Sammallahti 1988: 537)
- PU **muja-* ‘to become happy; happiness; smile’ > PSam. **mâjâ-* (Aikio 2002: 22)
- PU **luč’a* ‘wooden tool’ > PSam. **jâcâ* (Aikio 2002: 12–13; Aikio 2006: 29)
- PU **puna-* ‘weave’ > PSam. **pân-* (Sammallahti 1988: 539)

In a number of examples, like PU **kuma-* > PSam. **kâmâ-* ‘fall over’ and PU **muja-* > PSam. **mâjâ-* ‘become happy; smile’, the final PU **-a* was eventually changed into PSam. **-â*, evidently only after it had caused the reduction of **u*. The exceptions, in which **u* was not reduced to **â* before a second-syllable PU **-a*, all contained an intervocalic **-w-* in Proto-Uralic:

- PU **kuwakka* ‘long’ > PSam. **kuâkâ-nâ* ‘long ago’ (Aikio 2012: 244)
- PU **puwa-li-* ‘swell’ > PSam. **puâl-* (Aikio 2012: 244)
- PU **puwa-* ‘blow’ > PSam. **puâ-* (Aikio 2012: 244)
- PU **(w)uwa-* ‘current; flow’ > PSam. **wuâ* ‘current’ (Aikio 2012: 244–245)

The second rule mentioned by Sammallahti, whereby **u* changed to **â* when a following **-i* was lost and the intervening consonant was neither a lateral nor **x*, is less certain. The cases where **u* should have been retained due to the second-syllable vowel **-i* instead of **-a*, but was reduced instead, are:

- PU **kuñili* ‘tear’ > PSam. **kâñâlâ* (Sammallahti 1988: 537)
- PU **kudi* ‘morning’ > PSam. **kâr* ‘id.’ (Aikio 2002: 42–43)
- PU **kulki-* ‘run, flow’ > PSam. **kâj-* (UEW: 198; Sammallahti 1988: 544)
- PU **suñi* ‘summer’ > PSam. **tâñ* (Sammallahti 1988: 540)
- PU **nusi-* ‘scratch’ > PSam. **nât-* (Sammallahti 1988: 538)

Of these, Sammallahti’s additional rule is able to account for PU **suñi* ‘summer’ > PSam. **tâñ*, PU **kudi* ‘morning’ > PSam. **kâr* and PU **nusi-* ‘scratch’ > PSam. **nât-*. Assuming an additional rule before a lateral or **x* are the developments of PU **tuli* ‘fire’ > PSam. **tuj*, PU **ulki* ‘shaft’ > PSam. **uj*, PU **suxi-* ‘row’ > PSam. **tu(â)-* (cf. the list below). It does not, however, account for the development of **kulki-* ‘run, flow’ > PSam. **kâj-*, for which no explanation is available. On the other hand, PU **kuñili* ‘tear’ might have been influenced by the verb PU **kuña-* ‘close eyes’ > PSam. **kâñâ-* (Pystynen 2014).

The following list contains those words in which PU **u* did not change to **â*, apart from the words with intervening PU **w* already given above. It is indeed clear that all remaining examples have a second-syllable PU **-i*, and not **-a*.

- PU **purki* ‘blizzard, smoke; swirl, spray’ > PSam. **pur* ‘id.’ (Aikio 2002: 25–27)
- PU **kuñici* ‘urine’ > PSam. **kunsâ-* (Sammallahti 1988: 537)
- PU **pučki* ‘tube’ > PSam. **pucâ* (Janhunen 1981: 236; Aikio 2014a: 11–14)
- PU **suxi* ‘ski’ > PSam. **tutâ* (Sammallahti 1988: 540)
- PU **suxi-* ‘row’ > PSam. **tu(â)-* (Sammallahti 1988: 540)
- PU **tuli* ‘fire’ > PSam. **tuj* (Sammallahti 1988: 540)
- PU **tumti-* ‘know’ > PSam. **tumtâ-* (Sammallahti 1988: 541)
- PU **?uji-* ‘swim’ > PSam. **uâ-* (UEW: 542; Sammallahti 1988: 536 PU **uxi-*; Aikio 2020: 82–83 PU **?owji-*)
- PU **ulki* ‘pole’ > PSam. **uj* (Sammallahti 1988: 536; cf. Aikio 2002: 26)
- PU **ukti* ‘passage, way’ > PSam. **ut(â)* (Aikio 2012: 230)
- PU **juri-* ‘spin; go round’ > PSam. **ju/ürâ-* ‘get lost; forget’ (Aikio 2002: 46–48)

Similar to the change from PU **u* to PSam. **â*, PU **i* was reduced to PSam. **â* in some cases, although the circumstances have been less clearly defined. Janhunen considers a following lateral to be one cause of this reduction, as

in PU **ćilmä* > PSam. **sajmä* ‘eye’ and PU **?pilmitä* > PSam. **pajmätä* ‘dark’, but he remarks that there are other exceptional cases (Janhunen 1981: 225; 237; 253–254). Based on examples like PU **ipsi-* ‘smell’ > PSam. **aptä-*, PU **itä-* ‘appear’ > PSam. **ätä-*, and PU **imi-* ‘suck’ > PSam. **əm-* (in der. **ämmä-* ‘suckle, breastfeed’), the word-initial position of PU **i* was considered as a factor for reduction to PSam. **ə* (Aikio 2002: 24). Janhunen later pointed out that the Proto-Samoyed pronouns **mən-* ‘1sg.’ and **tən-* ‘2sg.’ take suffixes with front vowel harmony, so that these forms can be interpreted as containing a reduced **i* in PU **minä* and **tinä*, rather than a reduced PU **u* (Janhunen 2013: 214). The list of examples showing PSam. **ə* for PU **i* is thus, to my knowledge, as follows:

- PU **ćilmä* ‘eye’ > PSam. **sajmä* (Sammallahti 1988: 540)
- PU **imi-* ‘suck’ > PSam. der. **ämmä-* ‘suckle, breastfeed’ (Sammallahti 1988: 536; Aikio 2002: 24; Aikio 2020: 59)
- PU **ipsä-* ‘smell’ > PSam. **aptä-* (Sammallahti 1988: 536; Aikio 2020: 60–61)
- PU **ipsi* ‘smell’ > PSam. **aptä* (Sammallahti 1988: 536; Aikio 2020)
- PU **itä-* ‘appear’ > PSam. **ätä-* (Helimski 1993, Aikio 2002: 24)
- PU **minä* ‘1sg.’ > PSam. **mən-* (Janhunen 2013: 214)
- PU **?pilmitä* ‘dark’ > PSam. **pajmätä* (Sammallahti 1988: 539)
- PU **tinä* ‘2sg.’ > PSam. **tən-* (Janhunen 2013: 214)

As regards the noun and verb for ‘smell’, it has been pointed out that there was a distinction between the noun PU **ipsi* and the verb **ipsä-*, with differing second-syllable vowels. The noun is reflected in NenT *ḡabt°* ‘smell’ and PSaa. **epsę* ‘id.’, while the verb in **-ä* underlies NenT *ḡabtve-* ‘smell (intr.)’ and PSaa. **epsē-* ‘id. (intr./tr.)’ (Pystynen 2014; Aikio 2020: 60–61). Thus, in PSam. **aptä* ‘smell’ the vowel may be analogical to the verb **ipsä-* > PSam. **aptä-*, and the vowel reduction of PU **imi-* might have occurred specifically in the derivation **ämmä-* > NenT *ḡamʷa-*, apparently the only reflex of this verb in Samoyed (see Aikio 2020: 59). Based on such considerations, Pystynen (2014) suggests that it might therefore be possible to define the reduction of PU **i* with the same rule as that for **u*, namely as a type of *a*-umlaut affecting high vowels. Of the examples where PU **i* was retained, the only instances in an *a*-stem are the derivations PU **pidkä* ‘high’ > PSam. **pirkä* and PU **ńimćä* ‘breast’ > PSam. **ńimsä*, which may have been influenced by their underived counterparts, PSam. **pir* and PSam. **ńim-* respectively.

- PU **pidi* ‘high’ > PSam. **pir* (Sammallahti 1988: 539)
- PU **pidkä* ‘high’ > PSam. **pirkä* (Janhunen 1981: 239)
- PU **nimi* ‘name’ > PSam. **nim* (Sammallahti 1988: 538)
- PU **ńimi-* ‘suck’ > PSam. **ńim-* (Sammallahti 1988: 536; Aikio 2002: 24)
- PU **ńimćä* ‘breast’ > PSam. **ńimsä* (Aikio 2002: 23–26)

It thus seems reasonable to assume parallel developments of PU **i* to PSam. **ə* and PU **u* to PSam. **ə̂* before a second-syllable PU **-A*, as argued by Pystynen (2014). At least in the case of **u* this was blocked or reversed by an intervening intervocalic **w* (but not by a pre-consonantal **w* on account of PU **suwda* ‘finger’ > PSam. **təjā*). This change can be dated before the reduction of some PU **-A* to PSam. **-ə*. It should also be dated before PU **sewimä* > PSam. **timä* ‘tooth’ had an **i* in the first syllable (whatever the exact development; cf. section 2 sub 3), as we should otherwise expect PSam. **tämä*.

In both developments it may be assumed that the initial result of the change was not a simple **[ə]*. Words containing the vowels **i* and **u* remained faithfully integrated in their original harmonic class, which prompted Helimski to suggest that they might still have been distinct in Proto-Samoyed (Helimski 1993). Starting from original **i* and **u*, an intermediary stage like **[ɪ]* and **[ʊ]* could be posited (cf. Pystynen 2014), before those two eventually became further reduced to **ə* and **ə̂*. At what point exactly the two reduced vowel phonemes did fully merge phonetically is difficult to determine.

3.3 The development of PU **ü*

PU **ü* is another vowel that changed in Samoyed, so that we need to take its developments into account for our reconstruction of the pre-Proto-Samoyed vowel system. Although both Proto-Uralic and Proto-Samoyed are reconstructed with a front rounded vowel **ü*, the development of PU **ü* to PSam. *i* and its relation to PSam. **ü* is not straightforward. The two vowels often do not correspond, so that PSam. **ü* is usually not inherited from PU **ü*. Instead, PU **ü* was unrounded to **i* in Samoyed (Janhunen 1981: 247). A new PSam. **ü* arose both from secondary rounding of a Proto-Uralic unrounded vowel (e.g., PU **kävdi* ‘rope’ > PSam. **kürə*; Aikio 2006: 19–20) and from loanwords (e.g., **jür* ‘100’ from Turkic **yür*; Janhunen 1998: 477). I list here the examples that I am aware of:

- PU **dümä* ‘lime’ > PSam. **jimä* (Sammallahti 1988: 537)
- PU **üwä* ‘belt’ > PSam. **n/jiä* (Sammallahti 1988: 536; Aikio 2012: 230)
- PU **tütki*- ‘open, spread out’ > PSam. **titə*- (Aikio 2006: 26)
- PU **ćüdi* ‘charcoal’ > PSam. **sijə* (Sammallahti 1988: 540)
- PU **nüdi* ‘shaft’ > PSam. **nir* (Sammallahti 1988: 538)
- PU **süli* ‘fathom’ > PSam. **tij* (Sammallahti 1988: 540)
- PU **üli*- ‘space above’ > PSam. **i-* (Sammallahti 1988: 536)

There are some examples that seem to have retained PU **ü*, however. According to Janhunen (1981: 247, 254–255), the **ü* in these Proto-Samoyed words should be interpreted as secondary, so that the regular outcome was really PSam. **i*.

- PU **künti* ‘fog’ > PSam. **küntə* (Sammallahti 1988: 537)

- PU **ćünćä* ‘breast’ > PSam. **sünsə* (Sammallahti 1988: 540)
- PU **ńüktä-* ‘pull’ > PSam. **ńüt-* (Pystynen 2017)
- PU **kütki-* ‘tie’ > PSam. **küt-* (Pystynen 2017)

It is as yet uncertain how these words could have retained (or regained) their old vowel **ü*. Basing himself on PU **künti* ‘fog’ > PSam. **küntə* and PU **ćünćä* ‘breast’ > PSam. **sünsə*, Janhunen considered sporadic rounding of **i* to **ü*, retention of **ü* in a closed syllable, retention before a nasal, and areal effects from Ugric (Janhunen 1981: 254–255). Pystynen (2017) suggests that in **künti* ‘fog’, as well as in **ńüktä-* ‘pull’ and **kütki-* ‘tie’ (his proposed etymologies), the **k* could have been labialised before **ü* was unrounded (i.e., **künti* > **k^winti*, etc.). This new **k^w* could then later relabialise the neighbouring **i* (**k^winti* > **künti*, etc.). This explanation does not work for **ćünćä* ‘breast’, however, as it does not contain a velar; and on account of **ćüd’i* > PSam. **sijə* ‘charcoal’, a preceding PU **ć-* probably cannot be regarded as crucial for preserving or re-rounding **ü* either.

Thus, the history of **ü* in Samoyed has not yet been entirely elucidated. Nevertheless, the general developments can be added to the relative chronology. Firstly, it appears that PU **ü* only turned into **i* after original PU **i* was lowered to **ə* before **-A*, although the only secure etymology supporting this is PU **d’ümä* ‘lime’ > PSam. **jimä* instead of PSam. ***jämä* (Pystynen 2014). Secondly, PU **ü* must have changed into **i* before pre-consonantal **w* was lost or merged with the preceding vowel, yielding a new **ü*. This **ü* was not unrounded and remained in Proto-Samoyed; e.g., **käwdi* ‘rope’ and **jāwji* ‘lichen on trees’ > PSam. **kürə* and **jūjə* respectively (Aikio 2006: 13–14).

3.4 PU **a* and PSam. **ā*, **a*

The Proto-Uralic vowel **a* can be reflected in Samoyed as **ā* or **a*, which means that the Proto-Samoyed vowel inventory has gained an extra low vowel. This split of PU **a* into PSam. **ā* and **a* is another aspect of Samoyed historical phonology that is not yet entirely understood. The standard development seems to have been PU **a* to PSam. **ā*, as this reflex is represented by the majority of examples. Aikio (2002: 40) takes PSam. **a* as the regular reflex of PU **a* before **jC* or **ćC*, as exemplified by PU **aćkali-* ‘step’ > PSam. **asāl-*, PU **kajwa* ‘spade’ > PSam. **kajwā* and PU **wajŋi* ‘breath; spirit’ > PSam. **wajŋ*. He furthermore observes that this change should have occurred before **a* and **o* merged as **ā* in most environments, since, e.g., PU **moćki-* ‘wash’ became PSam. **māsə-*, and not ***masə-*. As argued by Aikio (ibid.: 41), the development of PU **aćkali-* > PSam. **asāl-* shows that this **a* became a separate phoneme before **k* was lost in clusters with an obstruent. However, there is an exception to this development in the form of PU **kajšo-* ‘be sick’ > PSam. **kājto-* (Aikio

2014a: 3–5), which does not show the change from PU **a* to PSam. **a* before **jC*. There are also a number of examples of PU **a* > PSam. **a* that do not share the same conditioning, so that a further explanation is required. Zhivlov has identified the reduction of certain second-syllable **-a* to **-â* as an environment in which PU **a* became PSam. **a* instead of **â* (Zhivlov 2014). This works for at least PU **kama* ‘skin, shell’, **taka* ‘behind’, **pata* ‘pot’, **aŋa-* ‘open’ in the list below, as well as for **ačkali-* ‘step’.

- PU **ačkali-* ‘step’ > PSam. **asâl-* (Aikio 2002: 40; Aikio 2020: 4)
- PU **taka* ‘behind (relational noun)’ > PSam. **tak(â)-* (UEW: 506–507; Aikio 2015b: 56)
- PU **kama* ‘skin, shell’ > PSam. **kamâ* (UEW: 120–121; Aikio 2015b: 55)
- PU **kajwa-* ‘dig’ > PSam. **kajwâ* (Aikio 2002: 41–42)
- PU **kari* ‘skin, bark’ > PSam. **kar* ‘skin, surface’ (Aikio 2012: 233)
- PU **jast* ‘chilly weather’ > PSam. **jat* (UEW: 637; Aikio 2015b: 52)
- PU **pata* ‘pot’ > PSam. **patâ-* ‘put something in a pot’ (UEW: 358; Aikio 2015b: 55)
- PU **aŋti* ‘blade’ > PSam. **aŋtâ* (Aikio 2015b: 52)
- PU **aŋi* ‘mouth’ > PSam. **aŋ* (Aikio 2015b: 65)
- PU **aŋa-* ‘open’ > PSam. der. **aŋâ-r-* (Aikio 2015b: 54)
- PU **čača-* ‘grow’ > PSam. (der.?) **caci-* (Aikio 2014a: 14–16)

The majority of cases show that PU **a* developed into PSam. **â*, merging with one of the reflexes of PU **o*. The two phonemes did not merge completely, however, because **o* was retained under some conditions. We might date the split of PU **a* into **â* and **a* as well as the split of **o* into **â* and **o* closely together. It looks as if the large majority of PU **a* became rounded at some stage and thereby encroached upon the territory of a low allophone of **o*. Since not every PU **a* was rounded and not every **o* came to be pronounced as a low vowel, the two vowel phonemes ended up yielding three. If the sequences **jC* and **-ćC-* indeed conditioned PU **a* to become PSam. **a* rather than **â*, the change must be dated before the simplification of clusters like PU **-ćk-* > PSam. **-s-* (Aikio 2002: 41).

3.5 *Relative chronology of Samoyed vowel developments in stages*

We may now put the changes discussed in an overview. Firstly, the split of PU **e* into high **i* and mid **e* seems to have been an early development in pre-Proto-Samoyed at least if we take the **-a* in the second syllable as one of the main conditioning factors. Since raising before a coda nasal is assumed to be another factor, it should probably have occurred before the loss of second-syllable **-i* as well, on account of the development of PU **d̥emi* ‘bird cherry’ to PSam. **jem* instead of to ***jim* (see section 3.1). The reduction of PU **u* and **i* in phonetic

TABLE 3 The development of the Samoyed vowel system from Proto-Uralic in four stages

PU			Pre-PSam. 1			Pre-PSam. 2			PSam.		
i	ü		i	ü	u	i (ü)	ï	u	i	ü	u
						ə		â		ə â	
e	ɛ	o	e	ɛ	o	e	ɛ	o	e	ɛ	o
ä	a		ä	a?	â	ä	a?	â	ä	a	â

terms is difficult to date, but phonologically it would have become significant at the latest when PU **ü* became unrounded to **i*, or when the two reduced vowels merged as one. The **i* from PU **ü* is not reduced to **ə*, even if it stands in a position where reduction of original **i* would be expected according to the conditioning posited by Pystynen (2014). Of course, new **ü* from earlier diphthongs only arose after old **ü* had disappeared (sections 3.2 and 3.3), since **ü* arising from such a source never becomes PSam. **i*. The phonologisation of PSam. **a* is difficult to date, because the exact conditions are to some extent still unclear. If clusters starting with a palatal consonant indeed induced a preceding PU **a* to remain **a*, as argued by Aikio (2002), this change must have occurred before **k* was lost when next to another obstruent (section 3.4).

Some external information on the Samoyed system may be gleaned from possible borrowings into (and perhaps also from) an early stage of Yukaghir. For our purposes it is interesting that Samoyed reduced vowels correspond to Yukaghir **a* (e.g., PU **pud'a*- ‘hit, split, break’ > PSam. **pâjâ*- > PYuk. **paj*- ‘strike, hit’), so that they should have been considerably lowered at the time of Samoyed-Yukaghir contact (see Aikio 2014c). Reliably dating this contact in absolute terms is not yet possible, however.

To conclude, the earliest changes that occurred in pre-Proto-Samoyed after Proto-Uralic seem to be the split of PU **ɛ* into **ɛ* and **ï*, and possibly the rise of “new” **a* next to **â* from PU **a* (and **o*). The resulting system is illustrated in the second column of Table 3. The following change to pre-Proto-Samoyed (2) in the third column would involve phonologisation of the reduced allophones of **i* and **u* as **ə* and **â*. PU **ü* may have been unrounded around the same time, but it does not seem to have been absent from the system for a particularly long time, and the reason for its apparent retention in some words is still obscure.

4 Development of the pre-Proto-Tocharian vowel system

Let us now move on to the Tocharian vowel system. As explained in section 2, the difficulties for Tocharian in Peyrot's comparison primarily center around the status of palatalisation in the phonological system at the relevant time depth. We shall therefore try to get a better view of that development in particular. In this section I will explore what the vowel system should have looked like before palatalisation occurred, based in part on an internal reconstruction of the Tocharian phonological system. I will mainly differentiate between on the one hand the Proto-Tocharian phonological stage, in which a large consonant system with palatalised consonants existed, and on the other hand an earlier stage of pre-Proto-Tocharian before palatalisation took place. At this earlier stage the vowel system was larger, because a number of vowel mergers that went hand in hand with palatalisation had not yet occurred. Since my goal is to offer a phonological reconstruction of the vowel developments from Proto-Indo-European to Proto-Tocharian, these mergers have to be undone in order to arrive at the older situation.

4.1 *Tocharian before palatalisation: internal reconstruction*

My understanding of palatalisation as a sound law is as follows: before phonemic palatalisation occurred, consonants were automatically pronounced as palatalised before a following front vowel (e.g., /Ci/ = [Ci]). At first this was a predictable, phoneme allophonic feature. However, when a merger occurred between front and non-front vowels, the palatalisation of the consonant became phonologically relevant. As a result, while a number of vowel phonemes merged, each consonant split into two variants, one palatalised and one non-palatalised or neutral.⁶ This can be schematically represented as /Ce/ vs. /Cë/ > /C^ye/ vs. /Ce/, where the vowel phonemes front /e/ and non-front /ë/ merge to yield new /e/. The phonological contrast of frontedness that was initially expressed on the vowels was transferred to the preceding consonant in the form of palatalisation. Such a change happened in the prehistory of Tocharian to yield the Proto-Tocharian system with palatalised consonants (represented as */C^y/). Thus, to retrieve the stage of pre-Proto-Tocharian right before palatalisation took place, we can imagine the reverse development, transferring the palatality of consonants back to the following vowel: a sequence */C^yV/ goes back to earlier */CV[+front]/, while */CV/ reflects earlier */CV[-front]/.

6 Except */r/, which apparently did not have a palatalised counterpart */r^yy

This will be the basis of my reconstruction of the pre-Proto-Tocharian vowel system that existed before palatalisation.

Leaving aside for the moment Proto-Tocharian syllables of the type */C^ya/ and */C^yo/, which arose due to umlaut after palatalisation had run its course, the following syllable types were present in the Tocharian inherited vocabulary:

- */C^yi/—PT **w^yikən* ‘20’ (**d^yih₁k^ynt*), PT **w^yike* ‘place, location’ (**μeikos*)
- */C^ye/—PT **w^yente* ‘wind’ (**h₂μeh₁ntos*), PT **m^yen^ye* ‘moon’ (**meh₁nēs*)
- */Ce/—PT **keme* ‘tooth’ (**gomb^hos*), PT **yəkwe* ‘horse’ (**h₁ek^yuos*)
- */Ca/—PT **pat^rer* ‘father’ (**ph₂tēr*), PT **k^yəna* ‘wife’ (**g^wenh₂*)
- */Cā/—PT **wāstə* ‘house’ (**μeh₂stu-*), PT **prāt^rer* ‘brother’ (**breh₂tēr*)⁷
- */Co/—PT **okso* ‘ox, cow’ (**uksōn*), PT **kl^yomo* ‘noble’ (**k^yleumōn*)⁸
- */C^yu/—PT **l^yuke* ‘light’ (**leukos*)
- */Cu/—PT **ku* ‘dog’ (**k^yuō*), PT **wu* ‘two’ (**d^yuoh₁*)
- */C^yə/—PT **l^yəke* ‘river’ (**tek^wos*), PT **l^yəmə* ‘lake’ (**lim^y*)
- */Cə/—PT **kānte* ‘100’ (**k^yntom*), PT **təpre* ‘high’ (**d^hub^hro-*)

The syllables PT */C^yi Ca Cā Co/ are the easiest to project back to a stage before palatalisation: we can simply reconstruct them as */Ci Ca Cā Co/; understandably, only the first of those yielded a palatalised consonant. */C^ye/ and */Ce/ can be reconstructed as */Ce/ and */Cē/ respectively (this distinction is included in Peyrot’s reconstruction of pre-Proto-Tocharian). Using internal reconstruction, the distinction between */C^yu/ and */Cu/ could be reconstructed as */Cū/ vs. */Cu/ in purely phonological terms; but we will return to this below. The vowel /ə/ can be split into two as well, with a front (palatalising) and a back (non-palatalising) variant. To write the front variant, the sign */ĩ/ will be used for now. Meanwhile, */ə/ may be used for the moment to represent the non-front counterpart to front */ĩ/, but these vowels will be further discussed below.

4.2 Tocharian before palatalisation: the Proto-Indo-European perspective

In general, the reconstructed distinctions based on simply undoing palatalisation accord very well with the known Proto-Indo-European sources for these

7 I operate with separate PT **ā* (from PIE **ā*) and PT **o* (from—some—PIE **ō* and umlaut). The former gave TA *a*, the latter TA *o*, while both yield *o* in Tocharian B. For a different view on the development of Tocharian *o*-vowels, pointed out to me by an anonymous reviewer, see Burlak & Itkin (2003). They reconstruct a single PT **o* for both **ā* and **o*, with conditioned developments causing a split in Tocharian A.

8 Reflexes of PIE **ō* as PT **o* are apparently only found in the second syllable; in the first syllable it was conditioned by umlaut.

syllable types. For instance, the distinction between pre-PT **/Ce/* and **/Cë/* reflects PIE **Cē* vs. **Co*, with a front and a back vowel respectively. Phonological **/Cü/* as opposed to **/Cu/* may be interpreted as reflecting PIE **Ceu* vs. at least secondary **Cū* from PIE **Cō#*. Whether the development of PIE **eu* to PT **yu* really happened via an intermediary stage that can be described phonetically as an **ü* (**/[y]/*) is difficult to ascertain, however. Alternatively, a sequence of developments like more or less **eu* > **ĭw* > **y_əw* = PT **/yu/* seems possible as well, and a reconstruction **ü* is at this stage only used as a phonological notation whereby the difference between the palatalising and non-palatalising vowels lies purely in the frontedness of the vowel. The palatalising */i/* derives from PIE **ī*/**ih₁* and **ei*, so in principle a reconstruction of **/i/* or **/īy/* for the stage before palatalisation is possible. We will return to this point below.

First, the precursors to PT **ə* require a more elaborate discussion. The front variant provisionally written as **/ĩ/* derives from both PIE **i* and **e*, for which no different phonological values can be decisively established on the basis of Tocharian.⁹ The non-palatalising schwa **/ə/* often derives from PIE **u*, so it could be written as **/ũ/* instead. However, non-palatalising PT **ə* can also derive from the prop vowel used to vocalise the syllabic resonants, i.e., PIE **R̥* > pre-PT **əR*. It may thus be preferable to set up a still earlier three-way distinction of **/ĩ/* vs. **/ũ/* vs. **/ə/*, all three of which eventually merged as PT **/ə/*.

Two types of umlaut involving **u* support the conclusion that the contrast between short PIE **u* and the other vowels **i* and **ə* was retained for a while before all eventually yielded reduced PT **ə*. The *u*-umlaut seen in TAB *or* ‘wood’ < **erũ* < PIE **doru* and in TA *okät*, TB *okt* ‘8’ < **ektu* < PIE **oktō* was triggered by both PIE short **u* and by earlier long **ū* < PIE final **ō*. This shows that short PIE

9 The view that PIE **i* palatalised less than PIE **e* is, in my opinion, not supported by a compelling number of secure examples, and it goes against phonological naturalness and the known typology of palatalisation (cf. e.g., Bateman 2011 and Bhat 1978). But see, e.g., Pinault (2008: 423), who excludes palatalisation of bilabials, labiovelars and **s* by PIE **i*. The development next to labiovelars (TB *wase* ‘poison’ < **wiso*, etc.) is to be viewed as a special change of the **i* caused by a **w* or **k^w* before palatalisation per Ringe (1996: 66). TB *skiyo* ‘shadow’ is the only proper example of non-palatalising **i*, and since **i* does palatalise alveolar (TB *lyam* ‘lake’ < **lim̥*, TA *-ñc* ‘verbal ending 3pl.prs.act.’ < **-nti*), the alternative is to set up an initial wave of palatalisation of just those alveolars, after which **i* was centralised, before **e* palatalised everything. I do not think *skiyo* ‘shadow’ can carry the weight of proof by itself, although I admit that it is not an altogether impossible scenario. If one were inclined to do so, one could also posit a sound law whereby **i* was centralised not only after **w*, but also after **k* to account for *skiyo*. Centralisation of **i* can in certain contexts have occurred before centralisation of **e*, as pointed out by an anonymous reviewer. I think that TB *nrai** ‘hell’ from Skt. *niraya-*, which was adduced by another anonymous reviewer, was certainly borrowed after palatalisation had run its course.

**u* was still rounded at the time.¹⁰ The same umlaut occurred in TB *šñor* ‘sinew’ < PT **sʸnʸorə* < **snʸerǔ* < **snēru* < **sneh_ɪru* < (metathesised) PIE **sneh_ɪur* (see Del Tomba 2021), so it can most economically be applied to both **e* < **ē* and **e* < **ĕ* < PIE **o* after palatalisation and the concomitant merger of **e* with **ĕ*. Also TA **šont* ‘road’ < PT **sʸontə* < **sentǔ* < PIE **seh_ɪntu-* (Hilmarsson 1986: 23–27) testifies to this. This umlaut was probably phonologised when short PIE **u* finally lost its rounding and became **ə*. Before that, when **u* or */ǔ/ was itself still rounded, the rounding of preceding **e* would have been phonetically motivated and likely automatic.

With *o*-umlaut, in turn, PIE **u* and **eu* were changed into PT **o* at least before a following **o* < PIE **ō*. This affected PIE **uksōn* > PT **okso* ‘ox, cow’ and PIE **k_leumōn* > PT **k_lʷomo* ‘noble’. The first example shows that this change affected short **u* (not **ə*, cf. **wālōnts* > PT **wəlo* ‘king’ > TA *wäl*, TB *walo*, not PT ***wolo* > TA ***wol*, TB **(*w*)*olo*), and the second example confirms that this umlaut happened after palatalisation had taken place. So it turns out that our */ǔ/ was still an independent phoneme even after palatalisation, and sufficiently strongly rounded to serve as a trigger for umlaut. This logically means that it was also an independent phoneme before palatalisation.

An additional indication that PIE **u* and **ə* should be kept separate from the vowel **ə* that arose from syllabic resonants may be found in the rounding that short **u* seems to have caused, if the reconstructed sequence of changes in PIE **k_wi-so* > **ku-so* > **kū-se* > PT **k_wə-se* is correct (see Ringe 1996: 66). Admittedly, the Proto-Tocharian labiovelar may also have been preserved as such, i.e., PIE **k_wi-so* > **k_wu-so* > PT **k_wə-se*. However, if Ringe is right, the effect of short **u* contrasts with that of the **ə* that arose from syllabic resonants, which did not labialise a preceding velar, e.g., PIE **k_ṛntom* ‘100’ > PT **kānte*, not ***k_wānte*.¹¹

10 Incidentally, this makes it impossible to date the change of PIE **u* to **ə* before PIE **o* to **ĕ*. This means that the motivation for unrounding and centralising PIE **o* cannot have been systemic pressure for centralised PIE */i e u/, since **u* was apparently not yet unrounded or centralised to */ə/ when PIE **o* became **ĕ* (or even **e*). Such a motivation for the unrounding and centralisation of PIE **o* was implied by Peyrot (2019: 82).

11 An anonymous reviewer has pointed out that there are a few etymologies thought to show a change **k_ṛ* to **k_wər* (via **kur*), viz., TAB *k_wär* ‘age, grow old’ < if from PIE **ǵ_ṛh₂-* (see Adams 2013: 255; see also Kim 1999: 155–156 for a discussion of three proposed etymologies for this root); TA *kursär*, TB *k_warsär* ‘league’ < PIE **k_ṛs-r-u-* (Adams 2013: 253), suggesting that the syllabic resonant **r* behaved differently from **ṛ* and **ṛ̥*. Since a reflex with simple **ə*, not **u*, is needed for words like PIE **k_ṛntom* ‘100’ > PT **kānte*, even if that analysis is correct, I simply refer to the discussion of these etymologies by Adams, as well as to Ringe (1991: 81–83) and Burlak (2000: 124) for a discussion of the issue.

TABLE 4 Undoing palatalisation

PT	Pre-pal. pre-PT	With diphthongs	PIE
Cʸi	Ci	Ciy	Cih ₁ , Cei
Cʸe	Ce	Ce	Cē, Ceh ₁
Ce	Cē	Cē	Co
Ca	Ca	Ca	CH, Ch ₂ e, Cō?
Cā	Cā	Cā	Ceh ₂
Co	Co	Co	Cō
Cʸu	Cü	Ciw	Ceu
Cu	Cu	Cuw	Cū (< -ō#)
Cʸə	Cĭ	Ci	Ci, Ce
Cə	Cŭ, Cə	Cu, Cə	Cu, Cŕ

Based on all these considerations, taking both bottom-up and top-down reconstruction into account, the vowel system right before palatalisation occurred should contain at least the following simple vowel phonemes: **/i e a ā o u ě ĭ ŭ ə/*. PIE **ei* and **eu* could be written as **/i/* and **/ü/* respectively, but **/iy/* and **/iw/* would also be possible instead. If the former option is chosen, the system can be represented as **/i ü e a ā o u ě ĭ ŭ ə/*, that is to say, with the additional vowel phoneme **/ü/*. This is represented in the second column of Table 4. If an interpretation with diphthongs is chosen at this time depth, the system can be represented as **/iy iw e a ā o uw ě ĭ ŭ ə/* instead (keeping to the same order). Here the breve signs on **/ĭ/* and **/ŭ/* do not express a phonologically distinctive feature, so that they could also be dispensed with. This system is shown in the third column of Table 4. Other diphthongs at this stage would have been **/ay aw ěy ěw/* < **/ay aw oy ow/*, under either interpretation. Table 4 shows the different stages of phonological analysis and reconstruction just discussed.

4.3 The development of PIE **ō*

Now it is time to address the development of PIE **ō* in a bit more detail. The development of this vowel is one of the most contentious and difficult aspects of the history of the Tocharian vowel system. It is variously said to become PT **u*, **o*, **a* and even **ə*, in part depending on the context. The first of these outcomes, **u*, is obtained in word-final position, as shown by the words PT **ku*¹²

12 An alternative account of PT **ku* would have the development of PIE **ō* go to PT **ə*

‘dog’ < PIE **kūō*, PT **wu* ‘2’ < **d̥uoh₁*, TA **ampu-k* < PT **antpu-* < PIE **h₂ntb^hoh₁* ‘both’. The same change occurred in PIE **oktō* ‘8’, although it is not obvious from the base numeral TB *okt*, TA *okät*. In this word the final *-u* was apparently removed due to analogy with PT **šəptə* ‘7’ and PT **ňəwə* ‘9’, but it was preserved in the Tocharian A decad *oktuk* ‘80’ (Ringe 1996: 89–90, Kim 2018: 101). It is also seen in the vowel *o* in both TA *okät* and TB *okt*, which arose from **e* (PIE **o*) by *u*-umlaut. Somewhat more controversial is the application of this change to the first person ending PIE **-ō* (**-oH*), which could have yielded PT 1sg. **-u* (Ringe 1996: 89).¹³

The examples where a development of PIE **ō* > PT **o* seems absolutely necessary are PT **wəlo* ‘king’ < quasi-PIE **ulHōnts*, **oksō* ‘ox’ < PIE **uksōn*, and *mo*-adjectives like PT **klʷomo* ‘noble’ < PIE **kleumōn* (Pinault 2008: 421–422). We have to assume that these PT **o* in the second syllable represent a direct reflex of PIE **ō*. The raising to PT **-u* in final position (as in PIE **oktō* > pre-PT **oktū* ‘8’, PIE **kūō* > pre-PT **k(w)ū* ‘dog’) would have been hindered by the presence of final **-n(ts)* in these forms before these consonants were lost by apocope.

A number of examples have been interpreted as reflecting a development of PIE **ō* to PT **a*, but the exact reconstructions are almost all problematic in some way. Only examples in PIE **-ōr* > PT **-ar* seem to be reliable, and difficult to explain differently, namely PIE **h₁itōr* > PT **yatar* ‘road’, PIE **h₁imōr* > TA *ymār* ‘quickly’, PIE **uesōr* > PT **wʷəsar* ‘spring’; see Del Tomba 2020: 151 fn. 227 with references.¹⁴ This may point to a special development before final **r*, especially since the change from **ō* to **a* would be a surprising development, on account of the (more secure) opposite change of earlier **ā* to PT **ā* to TB *o*. A crossing of these vowels is difficult to understand. Ringe (1996: 93) tentatively leads **ō* through an intermediate stage **ē*, parallel to the development of short **o* > **ē*, which would have been lowered to merge with original **a* after long **ā* had been rounded to **ā*, after which it was raised to *o*. Theoretically, this seems possible, although there is no evidence for an intermediary stage **ē*.

instead, via **kūwū* < **kauwō* < **kūō*, with the raising conditioned by the preceding **w* (see Adams 2013: 190), but in any case, the normal word-final development of PIE **ō* to PT **u*, not **ə*, is confirmed by TA *oktuk* where a regular development of **ū* to schwa should have yielded ***oktāk*.

13 For other explanations for this ending, see Malzahn 2010: 29, with references.

14 One of the stronger remaining arguments for PIE **ō* > PT **a* is PT **aknatsa* ‘foolish’ next to Lat. *ignōtus* ‘ignorant; unknown’ and Skt. *ājñāta-* ‘unknown’ < PIE **ǵneh₃-to-*. However, PT **aknatsa* should now be seen as an originally feminine *nt*-participle **ǵ-nh₃-nt-ih₂*, with zero grade in the root per Friis (2021). This means that the PT **a* is a reflex of the laryngeal **h₃*, and not of long **ō*.

In the relative chronology it would have to be dated before the loss of vowel length (*ibid.*).

If a general change from **ō* to **a* in non-final position is correct, one might wonder if length was really still a factor at the relevant time, however. The relative chronology could be set up in such a way as to allow for already shortened **o* to become **a* when it did not occur in word-final position (as in PT **wəlo*, **okso*, **klʷomo* after apocope of **-n(ts)*). If such a sound change followed palatalisation, a newly shortened **o* should not interfere with the **e* derived via **ē* from original short **o*, so that it could be safely unrounded and merge with **a* from the laryngeals and **h₂e*. Meanwhile an intermediate stage **ā* from earlier long **ā* could have retained its distinctive labialisation, to later yield *o* in Tocharian B.

4.4 *Phonological reconstruction of the vowel system from PIE to PT*

Based on this reconstruction, we arrive at the stages of the vowel system as displayed in Tables 5 and 6. These tables should provide us with a good point of departure for our comparison with pre-Proto-Samoyed. Table 6 gives the same basic developments as 5, but with the interpretation of **i* and **u* as diphthongs. In both tables, in the change from the first to second column, vowel length was lost. The development from the second to the third column involves the phonemicisation of palatalisation. After *u*-umlaut and *o*-umlaut, short PIE **u* was finally unrounded, and it merged with **i* as the Proto-Tocharian reduced vowel **ə*.

5 Comparing the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems

The pre-Proto-Tocharian system before phonemic palatalisation may now be compared with the vowel systems that we have reconstructed for pre-Proto-Samoyed. There are actually two different possible comparisons to be further explored. First, in subsection 5.1, I will give comparison (1) with a pre-Proto-Samoyed system **/i (ü) e ä a? ʔ o u i ɛ ə ɛ/*, which can easily be compared with the pre-Proto-Tocharian system before palatalisation given in Table 5. After addressing some remaining questions in subsection 5.2, I will then discuss comparison (2) in sections 5.3 and 5.4. This second comparison will center around an earlier pre-Proto-Samoyed system **/i ü e ä a? ʔ o u i ɛ/*, and is based on the pre-Proto-Tocharian system as represented in Table 6.

TABLE 5 The development of the Tocharian vowel system in four stages

Post-PIE			Pre-pal.			Post-pal.			PT		
ī eu		ū	i “ü”		u	y i		(y) u	i		u
i	(R)	u	ĩ	ə	ũ	(y) ə		ũ		ə	
ē e		o ō	e	ë	o	(y) e		o	e		o
	a ā			a	ā		a	ā		a	ā

TABLE 6 The development of the Tocharian vowel system if (pre-)Proto-Tocharian high vowels are interpreted as diphthongs

Post-PIE			Pre-pal.			Post-pal.			PT		
ī eu i		ū	iy iw		uw	y iy		(y) uw	əy~i		əw~u
ē e	(R)	u	i	ə	u	(y) i		u		ə	
		o ō	e	ë	o	(y) e		o	e		o
	a ā			a	ā		a	ā		a	ā

5.1 *Comparison (1)*

The first comparison departs from the phonological reconstruction of pre-Proto-Tocharian before palatalisation as */i ü e a ā o u ĩ ǣ ũ ə/ (cf. Table 5). This vowel system is structurally quite similar to the pre-Proto-Samoyed system reconstructed as */i (ü) e ä a? ā o u ĩ ǣ ə â/ (pre-PSam. 2 in Table 7). The most significant difference is that the pre-Proto-Samoyed system may have had more low vowels than pre-Proto-Tocharian, namely pre-PSam. */ä a? ā/ as compared to pre-PT */a ā/. Aside from that the two systems are remarkably similar. The two vowel systems are shown next to one another in the middle column of Table 7.

The mechanism by which the post-PIE system */i e a o u ī ē ā ō ū ə/ would be changed to pre-PT */i e a ā o u ĩ ǣ ũ ə/ can be understood relatively straightforwardly as loss of distinctive vowel length. Since the resulting system is very similar to the one labeled “pre-PSam. 2” in Table 7, such a pre-Proto-Samoyed system may have influenced the developments of the pre-Proto-Tocharian system. The post-PIE system had six short vowels and five long vowels, while the Proto-Uralic, pre-Proto-Samoyed, and Proto-Samoyed systems only ever had short vowels. However, making up for the lack of vowel length, pre-Proto-Samoyed had a greater number of distinct vowel qualities, up to eleven in the

TABLE 7 Different stages of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems compared

Post-PIE			Pre-PT			PT		
ī i	əR	u ū	i “ü”	ə	u	i		u
			ĩ		ũ		ə	
ē e		o ō	e	ë	o	e		o
	a ā			a	ǣ		a	ǣ

Pre-PSam. 1			Pre-PSam. 2			PSam.		
i ü	ĩ	u	i (ü)	ĩ	u	i ü	ĩ	u
			ə		â		ə~â	
e	ę	o	e	ę	o	e ö	ę	o
ä	a?	a	ä	a?	ǣ	ä	a	ǣ

pre-Proto-Samoyed system under consideration. Table 8 shows how the various post-PIE vowels could have changed to the pre-Proto-Tocharian vowel system so that it became more similar to a system of the type exhibited by pre-Proto-Samoyed.

The long vowels **ī ē ā ō ū* could be seen as corresponding to PSam. **/i e ǣ o u/*. The contrast between PIE **ī/* and **/i/* could have been maintained by way of a reduction of the latter, which possibly became similar to the pre-Proto-Samoyed front reduced vowel **ə*. PIE **e* may have merged with PIE **i* as pre-PT **ĩ* at this stage as well. The contrast between **ū* and **u* could develop in the same manner as the contrast between **ī* and *i*, yielding pre-PT **u* and **ũ*. Short **a* occupied a similar place in the system as either pre-PSam. **ǣ* or, if it existed already, **a*. There was no secondary o-type phoneme comparable to short **o*, so perhaps it was changed to become unrounded like PSam. **ę* in the new system, in the same way as argued by Peyrot. Peyrot’s equation of pre-Proto-Tocharian **ə* with pre-Proto-Samoyed **ĩ* may also be upheld, albeit using only **ə* from syllabic resonants, and not from short PIE **i*, **e*, and **u*, as in Peyrot’s comparison.

5.2 *Comparison (1): problems*

The vowel comparison just discussed appears quite exact: every vowel phoneme that can be reconstructed for pre-Proto-Tocharian before palatalisation

TABLE 8 Typological relations between the vowel phonemes of pre-Proto-Tocharian and pre-PSam.2 per comparison (1)

PIE	Pre-PT	Pre-PSam.2
i, e	> ĭ	ə
a	> a	ä/a?
o	> ě	ę
u	> ŭ	ê
ə	> ə	ĩ
ī, ei	> i	i
ē	> e	e
ā	> ǣ (> o)	ǣ
ō	> o	o
ū	> u	u
eu	> ü	ü

can be associated with a pre-Proto-Samoyed vowel, and the reduction of PIE **i*, **e*, and **u* that occurred in Tocharian could be regarded as a side effect of pre-Proto-Samoyed substrate influence. The way in which the changes occurred due to this putative substrate influence is quite understandable, and the anachronism with palatalisation is resolved in this way. However, there are still some remaining problems that come into sharper focus here. Firstly, the issues associated with *u*-umlaut and *o*-umlaut are not fully resolved, seeing as PIE **u* would have become a reduced vowel well before palatalisation occurred. This means that its function as a trigger of umlaut after palatalisation is still not fully accounted for.

The short high vowels also give some reason to pause and consider how this comparison is to be imagined at the interface between phonology and phonetics. Especially the comparison between pre-PT **ə* and pre-PSam. **ĩ* needs to be discussed. With respect to the Proto-Tocharian palatal : neutral pair **/Cʲə/* : **/Cə/*, the front **/ĩ/* should be seen as the palatal counterpart of **/ə/*. The rounded counterpart to these, **/ũ/*, remained rounded until after palatalisation on account of *u*-umlaut and *o*-umlaut (see subsection 4.2), so that the change at the point of palatalisation itself was just a merger of **/ĩ/* and **/ə/*. If the latter is compared with pre-PSam. **j*, representing **[i]* (or **[u]*), this is actually surprising. On the face of it, such a phoneme should rather be the non-palatal counterpart to pre-PT **/i/*, not to a pre-PT **/ĩ/* (*±*[ɪ]*). So how was pre-PT **/i/*

at the stage before palatalisation different from */ə/ (±*[i]) in more than just frontedness? To say that the latter was also a reduced vowel, unlike the proposed Samoyed counterpart, is possible, but ad hoc. We might rather like to interpret pre-PT */i/ as *[i], although that leaves us to find an alternative interpretation for pre-PT */i/.

An additional consideration is that the Proto-Tocharian phonemes */i/ and */u/—at least morphologically—behaved as |ʏəy| and |əw| respectively based on the restoration of these phonemes in zero-grade forms, e.g., *CuC : *CeuC > *CaC : *CəwC >> CəwC : CəwC (= /CuC : CʏuC/) on the basis of the type, e.g., *CaRC : *CeRC > CaRC : CəRC (Adams 1978: 447; Malzahn 2010: 24–25). This behaviour accords quite nicely with the origin of pre-PT */i/ and */u/ as diphthongs: for the most part they reflect PIE *ei and *eu respectively. However, in pre-Proto-Samoyed, *i and *u would rather have been simple vowels. This means that if we take a pronunciation like that of Samoyed as an intermediate, their eventual behaviour as diphthong-like elements in certain contexts in early Tocharian would seem to require an extra step along the lines of PIE *ei > *i > *ʏəy > PT *i rather than PIE *ei > *ʏəy > PT *i. It would probably be more economical to think of PIE *ei and *eu as diphthongs until a later stage in pre-Proto-Tocharian, before they eventually yielded PT *i and *u.

5.3 Comparison (2)

This brings us to the second comparison. Considering the difficulties surrounding the interpretation of pre-PT */i/, */i/ and */ə/, we can look outside the confines of the basic vowel system. There we can observe that Proto-Uralic and early Samoyed allowed for a number of closing diphthongs; in particular, vowels could be combined with a glide *j or *w. In Proto-Uralic, we find, e.g., *suwda ‘finger’ (> PSam. *təjā) with *uw, and *ij is reconstructed for, e.g., PU *nijni ‘bast’ and *čijči ‘tannin’ (Aikio 2021: 162–164, also with a brief discussion on this sound sequence in Uralic). Although reflexes of PU *ij are, to my knowledge, not found in words that are also attested in Samoyed, it can be assumed that this sequence existed at least in early forms of pre-Proto-Samoyed. It may also be reconstructed at an intermediate stage in the development of, e.g., PU *čilmä ‘eye’ > pre-PSam. *sijmä > PSam. *səjmä. Perhaps *ij and *uw can be compared to the earlier pre-Proto-Tocharian long high vowels *ī and *ū.¹⁵ That way, what I wrote in the first comparison as pre-PT */i/ and */i/ would rather represent *[ij] and *[i] respectively. Such an interpretation accords with the

15 As pointed out by an anonymous reviewer, Germanic words with long high vowels borrowed into Saami are rendered with such diphthongs; e.g., PSaa. *vijðēs > SaaS vijries ‘wide’ (Lehtiranta 1989: 148–149).

TABLE 9 The development of the Tocharian system if high vowels are interpreted as diphthongs. The interpretation of **ə* as **[i]* is here written as such

PIE		Pre-pal.		Post-pal.		PT	
ī i	u ū	iy iw	uw	yiy	(y)uw	əy	əw
əR		i	[i]	(y)i	u	ə	
ē e	o ō	e	ë	(y)e	o	e	o
a ā		a ǎ		a	ǎ	a	ǎ

morphophonological behaviour of these phonemes discussed at the end of the previous subsection, and it would explain why at the time of palatalisation in pre-Proto-Tocharian, **ə*/*~***[i]* was the back variant of **[i]*/*~**[i]*, and not of **[i]*/*~**[ij]*.

Similarly, the contrast between pre-PT **u* and **[ū]* would be **[uw]* vs. **[u]* at the time; and what I wrote in the first comparison as **[ū]* from PIE **eu*, based on just the difference in palatality vis-à-vis **u*, can then probably be rendered accordingly as **[iw]*. This **[iw]* may after palatalisation have become **[jiw]* > **[juw]* by assimilation, eventually yielding palatalising PT **u*. The interpretation of the development of the Tocharian vowel system according to this comparison is reproduced in Table 9 above.

Such a sequence of events also accords well with *u*-umlaut. At the time of palatalisation, a merger occurred between **i* and **ə* (**[i]*) to become the new **i*, distinct from **iy* deriving from **ī* and **ei*. This **i* merged with **u* to yield PT **ə* only after *u*-umlaut and *o*-umlaut had run their course. As a side effect of this change, **iy* and **uw* automatically became **[ij]* and **[iw]*. Updated with these points, the typological comparison with pre-Proto-Samoyed can be adapted, as shown in Table 10.

A later merger of PIE **i* and **u* in Tocharian (well after palatalisation) could perhaps also explain why the final **-u* of Old Iranian loanwords was reduced and eventually apocopated. For example, OIr. **dzainu-* → TB *tsain* ‘arrow’ with the plural *tsainwa* suggests that it was originally adopted as a *u*-stem noun, reflecting its origin as a *u*-stem noun in the donor language (i.e., **tsainu* > later PT **tsainə*; cf. Peyrot 2019: 83). The development of the **u* in this word to **ə* could, depending on the dating of that change, be the same development as the merger of original short **u* with **i* as PT **ə*.¹⁶ An interpretation of **ə* (**[i]*)

16 In other respects the stage of Tocharian that was in contact with Old Iranian seems to have been phonologically advanced, and close to Proto-Tocharian (see Bernard fthc.)

TABLE 10 Typological comparison between the vowel phonemes of pre-Proto-Tocharian and pre-PSam.1 per comparison (2)

PIE	Pre-pal. Pre-PT	Pre-PSam.1
i, e	> i	i
a	> a	ä
o	> ë	ę
u	> u	u
ə	> ĭ	ĭ
ī, ei	> iy	ij
ē	> e	e
ā	> ǎ	ǎ
ō	> o	o
ū	> uw	uw
eu	> iw	iw

next to **u* as **[iw]* (rather than simply **u* as **[u]*) may also make it more understandable that a loanword like buddha became **pətə* ($\pm$ **[piti]*), and *sukha* $\rightarrow$ **səkʷə* ($\pm$ **[sikʷi]*) ‘happiness’. With the short vowel **[u]* in the source, an actual diphthong **[iw]* might have been inappropriate as a replacement. At the same time, a high central unrounded short **[i]* should have been phonetically relatively close to a short **[u]*, especially in a labial environment (as in both PT **pətə* and (adapted) PT **səkʷə*), making that a fitting option for loan adaptation even after original short **[ū]* was unrounded in Tocharian. There are other possible interpretations, of course. These developments could in theory indicate a reluctance to borrow foreign short **u* as PT **[u]* < **ū* for many generations after the latter had stopped contrasting with original PIE **u*. Or perhaps there was long period of time in which the sound law reducing short **u* to **ə* was active. In any case, why new (though admittedly marginal) **u* from **ū* could not be used to represent foreign **u* may also be explained if **[u]* was phonetically actually a diphthong like **[iw]*.¹⁷

17 The diphthongal interpretation may furthermore shed a different light on *o*-umlaut as it occurred in PT **kʷomo* < **kleumōn*. If this word contained a vowel **ū* or **(y)u*, the change is somewhat different from the one found in **ūkso* > **okso*, as seen in the interpretation of the system in comparison (1): both **ū/u* and **ū* would be targeted in the same way to become **o*. But perhaps a development in the way stated in this second comparisn, of

As can be seen from Table 10, comparison (2) does not have a clear counterpart for pre-Proto-Samoyed **ü*. This is not necessarily a problem, since absolutely complete system convergence does not always occur in language contact situations; and perhaps there was simply no appropriate sequence of sounds in pre-Proto-Tocharian at the time that corresponded to this vowel.

5.4 *Comparison (2): an alternative interpretation*

One additional possibility can be considered, as a subtype of comparison (2) just discussed. It is possible that the PIE **i* and **e* were not both rendered by **i* as in Table 10, but that these two front vowels were kept separate. After all, Samoyed also had both vowels. In principle we could then take the merger of PIE **i* and **e* as a later, independent pre-Proto-Tocharian development (just preceding or coinciding with palatalisation), and compare PIE **e* with pre-PSam. **e*. The long PIE **ē* may then be compared with pre-PSam. **ä*, while Tocharian **a* from PIE laryngeals would be closer to the Samoyed **a* of unfortunately uncertain age. These correspondences can be connected with a possible borrowing from Tocharian into Samoyed: Tundra Nenets *mʷenʷyʷ* ‘full moon’ < quasi-PSam. $\pm$ **mäniŋjə*¹⁸ from post-PIE **mēnē* > PT **mʷenʷe* ‘moon’ (Blažek *apud* Napol’skikh 2001: 371–372). It is clear that the medial nasal was a plain nasal PSam. **n*, not palatal **ñ*, as that would have become a glide *y* in Nenets between vowels; cf., e.g., PSam. **jñä* ‘tame’ > NenT *ŋjyʷ* (SW: 25). This means that the borrowing should have taken place before Tocharian palatalisation had occurred, in line with our hypothesis so far. On the basis of this etymology, we might see pre-PSam. **ä* rather than **e* as closer to pre-PT **ē*.

The same conclusion might be tentatively supported by Tocharian B *yasa* ‘gold’ < PT **wʷasa* < pre-PT **wesa*, if this was borrowed from pre-Proto-Samoyed

the type **eu* → **iw* > palatalising **yiw* > intermediary **yiw* can be assumed, with a change **yiw* > **yo(w)* resulting from the umlaut. This way, the affected vowel is the same in both instances: **ukso* > **okso* and **klʷuwmō* > **klʷo(w)mō*. The disappearance of **w* in the latter would be a phonologically straightforward change, since it is surrounded by labial sounds. This could be seen as an alternative to either the opposing sound changes in **ūkso* next to **klümo* or **ūkso* next to **klʷumo*. It is unfortunate that there are so few examples of these sound developments to attest to the exact changes that occurred, so that this suggested development must remain tentative.

- 18 In Tundra Nenets, *-(ʷ)yʷ* seems to be a suffix (cf. Salminen 1998: 170–172), so that there may originally have been a root PSam. **män(V)*. The Proto-Samoyed reconstruction as **mäniŋjə* (adapted from **meniŋjə* in Napol’skikh 2001) can probably only be considered an approximation, since non-initial **ü* normally becomes *ə* in Nenets (see Salminen 2012). Perhaps **mänäwja* would be better.

wāsa* ‘metal’ (Janhunen 1983, Kallio 2004, Peyrot 2019: 101). This Samoyed word is related to the likes of PSaa. **veaškē*, Fi. *vaski*.¹⁹ The cluster in the western Uralic languages shows that the original word should have had PU *-*čk*-. This was regularly simplified in Samoyed (cf. PU **ačkali*- ‘step’ > PSam. **asël*-), so that that branch provides the most plausible source for the Tocharian word (in which the **k* would not have been lost by sound law if it was adopted as *weska*). If the rise of distinct PSam. **a* is indeed connected with the position before palatal clusters like *-*čk*- (cf. 3.4), and the simplification happened before contact with Tocharian on account of this loan etymology, pre-Proto-Samoyed at that time did have a contrast */ä a ä/.²⁰

This does not change much in comparison (2) as already discussed, but aside from fitting in the ‘moon’ loan etymology, it also accords with the idea that PT **e* could be phonetically closer to *[ɛ] (although the value *[e] is not exactly excluded, cf. Pinault 2008: 420–421 with a brief discussion; see also Bernard fthc. on the hypothesis that this vowel was closer to *[æ]). Furthermore, the fact that PT **e* could beumlauted to **a* when followed by an **a* in the next syllable implies that it was probably not a particularly high vowel. The Samoyed **e*, on the other hand, was probably relatively high, as suggested by its later merger with **i* in most Samoyed languages, so this vowel might not be as good a comparandum for PT **e* on that account. The lower vowel PSam. **ä* perhaps fits better. The vowel replacement and correspondences involved in this scenario are shown in Table 11. In Table 12 the changes of the Tocharian vowel systems are shown again with the possible longer retention of PIE **e*, and later merger with **i*.

19 The vowel correspondences are difficult. In Samoyed, the *-*a* of the second syllable has not been explained. Aikio (2015b: 43) would rather take this as a kind of Wanderwort that has entered various Uralic languages independently, although I would think that the loss of **k* in Samoyed is suggestive of early adoption into that branch, even if it is not Proto-Uralic proper.

20 There remains another potential phonological problem with this etymology, as pointed out by an anonymous reviewer. Namely, the development of the PU (probably) affricate **č* to a sibilant in Samoyed might have to be dated later, even after Proto-Samoyed. As far as I am aware, this argument is based not on intervocalic reflexes of PU **č*, but only on the Tundra Nenets reflex of PSam. **s* in postconsonantal position as *c* ([ts]), and the merger of initial PSam. **s* + front vowel with **k* + front vowel as *k*- in Mator (Zhivlov 2018). If the correct reconstruction of the word is **wāca* rather than **wāsa*, it is indeed doubtful that this word is the origin of pre-PT **wesa*, but earlier deaffrication, especially in intervocalic position, is not excluded.

TABLE 11 Typological comparison of pre-Proto-Tocharian based on pre-PSam.1, with diphthongs and retention of Tocharian *e

PIE	Pre-PT	Pre-PSam.1
i	> i	i
e	> e	e
a	> a	a
o	> ĕ	ĕ
u	> u	u
ə	> i	ĩ
ē	> æ	ä
ā	> ǻ (> TB o)	ǻ
ō	> o (> a?)	o
ī	> iy	ij
ei	> ey	ej
ū	> uw	uw
eu	> ew	ew

6 Discussion

In the previous section I explored two possible comparisons between pre-Proto-Tocharian and pre-Proto-Samoyed, based on the relative chronology of vowel developments in the two languages. The two comparisons involve different stages of pre-Proto-Samoyed and different interpretations of Tocharian phonology around the time of palatalisation. I will now discuss which one is in my opinion more likely to reflect an actual instance of prehistoric substrate interference. I will then consider the time depth of the possible language contact.

6.1 *The two comparisons compared*

In comparison (1) I took the later pre-Proto-Samoyed system */i ü e ä a? ǻ o u ĭ ĕ ə â/ and compared it with a pre-Proto-Tocharian system */i ü e a ǻ o u ə ě ĭ ũ/. This looks like a close correspondence at first blush, but as discussed in section 5.2 there are a few remaining problems, even if it resolved some of the difficulties present in Peyrot’s (2019) comparison. This first comparison directly introduces the reduced antecedents of eventual PT *ə into the system in the form of */ĩ ũ ə/.

TABLE 12 The development of the Tocharian vowel system with longer retention of Tocharian *e

PIE		Pre-pal.			Post-pal.			PT	
ī i	u ū	iy iw		uw	yiy	(y)uw		yəy	(y)əw
Ṛ		i	i	u	(y)i	u		ə	
ē e	o ō	e	ë	o	(y)e	o		æ	o
a ā		æ	a	ǎ	(y)æ	a	ǎ	a	ǎ

Palatalisation can then be applied rather straightforwardly to arrive at the Proto-Tocharian situation. On the Samoyed side, the comparison depends on a system that is fundamentally very similar to that of eventual Proto-Samoyed. The vowels that would have to have merged in Tocharian as a result of the pre-Proto-Samoyed substrate would be PIE *i and *e, and the only additional assumption needed is that the reduced PU *i and *u had not yet merged phonetically, and that that they had not yet been significantly lowered at the time.

Contacts between pre-Proto-Samoyed and pre-Proto-Yukaghir might testify that pre-Proto-Samoyed */ə/ and */â/ were relatively close to a low vowel *[a] at an early stage, since possible borrowings containing that Samoyed vowel show Proto-Yukaghir *a. In theory, contact between pre-Proto-Samoyed and pre-Proto-Yukaghir might have taken place later than contact between pre-Proto-Samoyed and pre-Proto-Tocharian, but roundedness of PU *u > reduced PSam. *â at the relevant time depth for contact with pre-Proto-Tocharian remains conjectural.

Comparison (2), between pre-Proto-Samoyed */i ü e ä a? ǎ o u i e/ and pre-Proto-Tocharian */i e a ǎ o u ə ë/, may provide an adequate solution to some of the remaining problems present in comparison (1). In comparison (2), earlier PIE *ī, *ei, *ū, and *eu should be interpreted as the diphthongs *iy, *uw, and *iw respectively in pre-Proto-Tocharian. A difficulty is that the opposition of Proto-Samoyed */ä a ǎ/ cannot (yet) be reliably established to any particular time depth, even though it accords well with this comparison. However, there is an indication that pre-PSam. *a could already have developed at the time of contact, if we accept the proposed borrowing of pre-PT *wesa from pre-PSam. *wäsa. In this word, a cluster PU *čk was reduced to *č > PSam. *s, so that the sound law whereby PU *a became PSam. *a (not *ǎ) before a palatal consonant in a closed syllable, as stated by Aikio, should already have taken place. Within comparison (2), the pre-Proto-Samoyed */a/ could be given a place as well (see 5.4), which allows for a continued differentiation between PIE *i and

**e* after pre-Proto-Samoyed substrate influence, a continuation of PIE **eu* as **ew*, PIE **ei* as **ej*, etc. The vowel correspondence in the loan etymology for NenT *mʲenʲuy*^o ‘full moon’ < pre-PT **mēnē* ‘moon’ may also point in this direction.²¹

In comparison (1), the chronology of *u*-umlaut (type **eru* > **orə* ‘wood’) and *o*-umlaut (type **ukso* > **okso*) combined with palatalisation is problematic, as is to some extent the relationship between */*ĩ*/ and */*ə*/, and the morphophonological behaviour of PT */*i*/ and */*u*/ as [ʲəy] and [əw]. These problems receive a more adequate solution in comparison (2). In this comparison we obtain a more direct phonetic comparison between pre-PSam. */*i*/ and pre-PT */*ə*/ as */*i*/ in the light of palatalisation. Among the high vowels and diphthongs, pre-Proto-Tocharian palatalisation would work as follows: */*Ci Ci Cu Ciy Cuw*/ > */*Cʲi Ci Cu Cʲiy Cuw*/. Eventual PT */*Cʲə Cə Cʲəy Cəw*/ could then easily be the result of a later merger of */*i*/ and */*u*/ as PT */*ə*/. This could additionally make the reduction of **u* in loanwords like OIr. **dzainu* → PT **tsainə* ‘arrow’ with plural **tsainwa* and the adoption of buddha as PT **pətə* more understandable. This second comparison makes no reference to the pre-Proto-Samoyed vowel */*ü*/, if that existed at the time (as it probably did).

The advantages and disadvantages of the two comparisons are summarised in Table 13. Both accord with the chronology of palatalisation in Tocharian, but only comparison (2) takes umlaut fully into account. The mechanism of the vowel changes is quite understandable in both comparisons: it involves a loss of phonemic vowel length without a great loss of distinctions, resulting in more vowel qualities. A difference is that in comparison (1) the Proto-Tocharian reduced vowel is immediately introduced as such, while in comparison (2) it is left as a later development. With the introduction of reduced vowels, the system in comparison (1) is more similar to Proto-Tocharian than the one in comparison (2). This means that fewer Tocharian vowel changes would be required to take place after the putative contact period in the case of comparison (1). However, under the assumption that the reduced vowel arose later, as in comparison (2), the behaviour of the high vowels pre-PT **i* and **ə* in palatalisation can be better understood. The PT **ə* found in loanwords with original **u* in the donor language also accords with a late shift from PIE **u* to PT **ə*, well after any contact with pre-Proto-Samoyed. Likewise, the morphophonological behaviour of PT **i* and **u* as diphthongs is in accordance with comparison (2), whereas it does not follow from comparison (1).

21 In that case, the merger of **e* with **i* in post-contact Tocharian would be similar to what happened in all Samoyed languages except Nganasan.

TABLE 13 Overview of which points are addressed by the two comparisons advanced in this paper

Point addressed	Comp. (1)	Comp. (2)
Chronology with palatalisation	+	+
Chronology with <i>u</i> -umlaut and <i>o</i> -umlaut	–	+
Mechanism for change PIE > pre-pal. pre-PT	+	+
Origin of Tocharian reduced vowel	+	–
Relation pre-PT * <i>ĩ</i> to * <i>ə</i> : pre-PSam. * <i>ĭ</i>	–	+
Behaviour of PT * <i>i</i> and * <i>u</i> as <i>y</i> əy, (<i>y</i>)əw	–	+

6.2 *Time depth, archaeology and genetics*

An important difference between the two comparisons is the time depth at which the contact between pre-Proto-Tocharian and pre-Proto-Samoyed should have taken place. In comparison (1) a greater number of specifically Samoyed vowel developments are required, which means that pre-Proto-Samoyed would have needed time to undergo those developments separate from the other Uralic languages. In comparison (2), on the other hand, the stage of the pre-Proto-Samoyed vowel system that is assumed is not as overtly Samoyed in nature. The only specifically Samoyed feature would be the split of PU **ę* into **ĭ* and **ę*, and perhaps the rise of a three-way distinction between the low vowels */*ä* a *ǣ*/ from an earlier Proto-Uralic two-way distinction. Aside from that, the system is very much the same as the one that is reconstructed for Proto-Uralic. Based on comparison (2), the contact could therefore be dated earlier on in the development of pre-Proto-Samoyed, closer to the break-up of Proto-Uralic. Additionally, in Tocharian slightly more changes are needed after contact in comparison (2), such as the reduction of **i*, **e* and **u*.

Based on archaeological considerations, the ancestors of the Tocharians may be associated with the Afanasievo Culture of ±3300–2500 BCE. If correct, this probably represented the northernmost location for pre-Proto-Tocharian, closest to the area where pre-Proto-Samoyed might have been spoken. Seeing as contact between pre-Proto-Tocharian and pre-Proto-Samoyed could be more or less geographically connected with the Afanasievo Culture, we might also expect a temporal connection. Proto-Uralic has in recent literature been estimated closer to 2000 BCE on account of contact with early Indo-Iranian around that time (Kallio 2006: 11–13, Parpola 2013: 162). Grünthal et al. (2022) estimate PU at around 2500 BCE on similar grounds, emphasising that contact with Indo-Iranian occurred after the individual Uralic branches had already

become separate entities. It would be strange if around this same time pre-Proto-Samoyed had already undergone numerous significant changes to the vowel system, however. This is because the other branches of Uralic had hardly begun to differentiate at the time of contact with Indo-Iranian in terms of their phonology, to the point that many loanwords appear to show regular correspondences as though they were inherited from Proto-Uralic. This means that comparison (1) would probably be chronologically more difficult than comparison (2).

Genetically, individuals from the Afanasievo Culture were almost indistinguishable from those associated with the Yamnaya Culture. Afanasievo genetic signatures also appeared in Dzungarian Basin south of the Afanasievo area towards the middle of the Afanasievo period, around 3000–2800 BCE. The individuals found in Dzungaria, while still predominantly of Afanasievo ancestry, also show admixture from local populations (Zhang, Ning, Scott et al. 2021). If there was an expansion from Afanasievo to Dzungaria along the west side of the Altai mountains, that also seems to be an area where contact with early pre-Proto-Samoyed could have occurred. The upper reaches of the Irtysh and Ob rivers flow from that part of Central Asia, and the area surrounding those rivers may have been inhabited by Uralic-speaking peoples at the time. To this day Ob-Ugric languages are spoken along those rivers, albeit further downstream, to the north-west. The ancestors of the Samoyeds eventually ended up by the middle Ob the middle Yenisei, farther to the east, but they might still have been on the way in that direction along the Irtysh flowing from the Altai around the time that the people from Afanasievo were in the process of migrating into Dzungaria.

Grünthal et al. associate the spread of early Uralic languages (after the split of Proto-Uralic) and early contacts between these languages and Indo-Iranian with the Seima-Turbino phenomenon. The Seima-Turbino phenomenon is also represented in sites on the Irtysh, Ob and Yenisei rivers (Grünthal et al. 2022; cf. their Map 1), as well as farther west surrounded by cultures associated with early Indo-Iranian. Seeing as the Tarim Basin mummies (2100–1700 BCE) show no signs of Afanasievo ancestry (Zhang, Ning, Scott et al. 2021), the arrival of Tocharian in the Tarim Basin may have been later than 1700 BCE. Until that time, the ancestors of the Tocharians could have remained in the Dzungarian Basin and the surrounding area immediately to the north, where the ancestors of the Samoyeds may also have migrated through on their way toward the middle Yenisei. At this stage I think that a contact scenario along these lines is most promising, although the specifics must remain tentative.

7 Summary and conclusions

I have argued in section 5 that the pre-Proto-Tocharian vowel system as it can be reconstructed right before palatalisation occurred (section 4) was typologically very similar to the vowel system that can be reconstructed for pre-Proto-Samoyed (section 3). The most significant Tocharian vowel developments are a shift from PIE **o* to eventual PT **e*, and the general loss of phonemic vowel length. While such developments are not unheard of, it is at least suspicious that these changes should result in a system that conforms to that of pre-Proto-Samoyed. I think that on account of the close possible match with pre-Proto-Samoyed, supported by the similarities in the consonant system, a comparison can be upheld.

The difficulties with the earlier vowel comparison by Peyrot (2019) were discussed in section 2: the relation to palatalisation in the consonant system, the mechanism of the change from post-PIE to pre-Proto-Tocharian under supposed influence from pre-Proto-Samoyed, and the relative chronology of the Samoyed sound changes. Addressing these difficulties formed the basis of the two comparisons adduced in this article. The pre-Proto-Tocharian vowel system used in comparison (1) allows for a quite direct derivation of the Proto-Tocharian system by means of palatalisation and a merger of three already reduced vowel phonemes into the single reduced vowel **ə* of Proto-Tocharian (see subsection 5.1). It provides an explanation for the various vowel developments in pre-Proto-Tocharian, as most of them can be framed as the result of influence of the pre-Proto-Samoyed substrate, especially the development of short vowels PIE **i*, **u* and **ə*.

However, some problems remain with this comparison (see subsection 5.2). These can mostly be solved under comparison (2), which posits pre-PT diphthongs **/ij uw/* (or perhaps still **/ej ew/*) and allows for a more reasonable chronology regarding the development of PIE **u* in relation to *u*-umlaut and *o*-umlaut, as well as possibly a change of late borrowed **u* to PT **ə*. The window for *u*-umlaut and *o*-umlaut to occur is larger in this chronology, and the morphophonological behaviour of PT **/i u/* as *|ʔəy əw|* would be more understandable (see subsections 5.3 and 5.4). Furthermore, since the pre-Proto-Samoyed stage assumed in comparison (2) is more archaic than the one used in comparison (1), it is easier to bring comparison (2) into line with the Afanasiev hypothesis of the ancestors of the Tocharians, or more generally a relatively early time of contact (see subsection 6.2). Based on these considerations, I prefer comparison (2) (see also subsection 6.1).

The fact that two different stages of pre-Proto-Samoyed can in principle be compared to different possible (phonetic) interpretations of the pre-Proto-

Tocharian vowel system can be seen as a side effect of the great time depth involved on both sides. This naturally creates a certain margin for uncertainty and interpretation. I hope that the relative chronologies outlined and summarised for the vowel developments in both pre-Proto-Tocharian and pre-Proto-Samoyed are nevertheless robust enough, and that they will prove useful for further research on these languages, although some difficult points inevitably remain to be solved in the future. Since in my reconstruction the consonant and vowel systems of pre-Proto-Tocharian and pre-Proto-Samoyed can be compared at the same time depth with respect to the status of palatalisation, this new comparison lends additional weight to the hypothesis that the ancestors of the Tocharians were in contact with the early Samoyeds.

Acknowledgements

I presented part of this paper at the online conference “Tocharian in Progress” on 10 December 2020, under the same title. I am thankful to those who commented after the presentation. I am furthermore grateful for comments made by Michaël Peyrot (Leiden), Sasha Lubotsky (Leiden), Mikhail Zhivlov (Moscow), and Sampsa Holopainen (Vienna) on earlier versions of this paper, and valuable discussions on the topic. I also thank the two anonymous reviewers. This research was financed by the European Research Council (ERC-2017-STG 758855).

References

- Adams, Douglas Q. 1978. Ablaut and umlaut in the Tocharian vowel system. *Journal of the American Oriental Society* 98(4): 446–450.
- Adams, Douglas Q. 2013. *A dictionary of Tocharian B*. Amsterdam & New York: Rodopi.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2002. New and old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57: 9–57.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2006. New and old Samoyed etymologies 2. *Finnisch-Ugrische Forschungen* 59: 9–34.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2012. On Finnic long vowels, Samoyed vowel sequences and Proto-Uralic *x. In *Per Urales ad Orientem: Iter polyphonicum multilingue. Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*, ed. Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi, & Erika Sandman, 227–250. Helsinki: Suomalais-ugrilainen Seura.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2013. Studies in Uralic etymology I: Saami etymologies. *Linguistica Uralica* 2013(3): 161–174.

- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014a. Studies in Uralic etymology II: Finnic etymologies. *Linguistica Uralica* 2014(1): 1–19.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014b. Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica* 2014(2): 81–93.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014c. The Uralic-Yukaghir lexical correspondences: genetic inheritance, lexical contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62: 7–76.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015a. Studies in Uralic etymology IV: Ob-Ugric etymologies. *Linguistica Uralica* 2015(1): 1–20.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015b. The Finnic ‘secondary e-stems’ and Proto-Uralic vocalism. *Journal de la Société Finno-Ougrienne* 95: 25–26.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2020. *Uralic etymological dictionary* (draft version of entries A–Ć). URL: https://www.academia.edu/41659514/URALIC_ETYMOLOGICAL_DICTIONARY_draft_version_of_entries_A_Ć_.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2021. Studies in Uralic etymology V: Permic etymologies. *Linguistica Uralica* 57(3): 161–179.
- Anthony, David W. 2007. *The horse, the wheel and language. How bronze-age riders from the Eurasian steppes shaped the modern world*. Princeton: Princeton University Press.
- Anthony, David W. 2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *Journal of language relationship / Вопросы языкового родства* 9: 1–21.
- Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and linguistics compass* 5(8): 588–602.
- Bednarczuk, Leszek. 2015. Non-Indo-European features of the Tocharian dialects. In *Words and dictionaries. A Festschrift for Professor Stanisław Stachowski on the occasion of his 85th birthday*, ed. Elżbieta Mańczak-Wohlfeld & Barbara Podolak, 55–67. Kraków: Jagiellonian University Press.
- Bernard, Chams. fthc. *Like dust on the Silk Road. An investigation of Iranian loanwords and of possible BMAC loanwords in Tocharian*. Ph.D. dissertation, Leiden University.
- Bhat, D.N.S. 1978. A general study of palatalization. In *Universals of human language*, ed. J.H. Greenberg, 47–92. Palo Alto, CA: Stanford University Press.
- Burlak, Svetlana A. 2000. *Istoričeskaja fonetika toxarskix jazykov*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Burlak, Svetlana A. & Ilya B. Itkin. 2003. A sound change that never happened: The fate of Proto-Tocharian *o (B o) in Tocharian A. *Tocharian and Indo-European studies* 10: 17–35.
- Del Tomba, Alessandro. 2020. *The Tocharian gender system. A diachronic study*. Ph.D. dissertation, Leiden University & Sapienza University of Rome.
- Del Tomba, Alessandro. 2021. Metathesis of PIE *-ur̥ in Tocharian. *Münchener Studien zur Sprachwissenschaft* 74.1: 51–85.

- Friis, Louise. 2021. Tocharian B agent nouns in *-ntsa* and their origin. *Indo-European linguistics* 9: 1–25.
- Grünthal, Riho, Volker Heyd, Sampsa Holopainen, Juha Janhunen, Olesya Khanina, Matti Miestamo, Johanna Nichols, Janne Saarikivi, & Kaius Sinnemäki. 2022. Drastic demographic events triggered the Uralic spread. *Diachronica* 39(4): 490–524.
- Hackstein, Olav. 2017. The phonology of Tocharian. In: *Handbook of comparative and historical Indo-European linguistics*, ed. Jared Klein, Brian Joseph, & Matthias Fritz, Vol. 2, 1304–1335. Berlin: De Gruyter.
- Hall, T.A. 2000. Typological generalizations concerning secondary palatalization. *Lingua* 110(1): 1–25.
- Helimski, Eugen (Евгений Хелимский). 1993. Prasamodijskije *â i *ä: praural'skije istočniki i nganasanskije refleksy. In *Hajdú Péter 70 éves*, ed. Marianne Bakró-Nagy & Enikő Szij, 125–133. Budapest.
- Helimski, Eugen (Евгений Хелимский). 2005. The 13th Proto-Samoyedic vowel. In *Mikola-konferencia 2004*, ed. Beáta Wagner-Nagy, 27–39. Szeged: SzTE Department of Finnougristics.
- Hilmarsson, Jörundur. 1986. *Studies in Tocharian phonology, morphology and etymology, with special emphasis on the o-vocalism*. Ph.D. dissertation, Leiden University.
- Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic. Critical overview of the sound substitutions and distribution criterion*. Ph.D. dissertation, University of Helsinki.
- Ivanov, Vyacheslav V. 1985. Tocharian and Ugrian. In *Studia Linguistica, Diachronica et Synchronica. Werner Winter sexagenario anno MCMLXXIII gratis animis ab eius collegis, amicis discipulisque oblata*, ed. Ursula Pieper & Gerhard Stickel, 411–419. Berlin & New York: De Gruyter.
- Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77: 219–274.
- Janhunen, Juha. 1983. On early Indo-European-Samoyed contacts. In *Symposium saeculare Societatis Fenno-Ugricae*, ed. Juha Janhunen, Anneli Peräniitty, & Seppo Suhonen, 23–42. Helsinki: Suomalais-Ugrilainen Seura.
- Janhunen, Juha. 1998. Samoyedic. In *The Uralic languages*, ed. Daniel Abondolo, 457–479. London & New York: Routledge.
- Janhunen, Juha. 2013. Personal pronouns in Core Altaic. In *Shared Grammaticalization. With special focus on the Transeurasian languages*, ed. Martine Robbeets & Hubert Cuyckens, 211–226. Amsterdam: John Benjamins.
- Kallio, Petri. 2001. Phonetic Uralisms in Indo-European? In *Early contacts between Uralic and Indo-European*, ed. Christian Carpelan, Asko Parpola, & Petteri Koskikallio, 221–234. Helsinki: Suomalais-Ugrilainen Seura.
- Kallio, Petri. 2004. Tocharian loanwords in Samoyed? In *Etymologie, Entlehnungen und Entwicklungen. Festschrift für Jorma Koivulehto zum 70. Geburtstag*, ed. Irma Hyvärinen, Petri Kallio, & Jarmo Korhonen, 129–137. Helsinki: Société Néophilologique.

- Kallio, Petri. 2006. Suomen kantakielten absoluuttista kronologiaa. *Virittäjä* 110: 2–25.
- Kim, Ronald I. 1999. The development of labiovelars in Tocharian: a closer look. *Tocharian and Indo-European studies* 8: 139–187.
- Kim, Ronald I. 2018. *The dual in Tocharian. From typology to 'Auslautgesetz'*. Dettelbach: J.H. Röll.
- Krause, Wolfgang. 1951. Zur Frage nach dem nichtindogermanischen Substrat des Tocharischen. *Zeitschrift für vergleichende Sprachforschung* 69: 185–203.
- Kroonen, Guus, Gojko Barjamović, & Michaël Peyrot. 2018. Linguistic supplement to Damgaard et al. 2018: *Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian*. DOI: <http://doi.org/10.5281/zenodo.1240524>.
- Kümmel, Martin Joachim. 2007. *Konsonantenwandel. Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.
- Lehtiranta, Juhani. 1989. *Yhteisaamelainen sanasto*. Helsinki: Suomalais-Ugrilainen Seura.
- Mallory, James P. 2015. *The problem of Tocharian origins. An archaeological perspective*. Philadelphia.
- Malzahn, Melanie. 2010. *The Tocharian verbal system*. Leiden: Brill.
- Mikola, Tibor. 2004. *Studien zur Geschichte der samojedischen Sprachen*. Szeged: SzTE Finnisch-Ugrisches Institut.
- Napol'skikh, Vladimir. 2001. Tocharisch-uralische Berührungen: Sprache und Archäologie. In *Early contacts between Uralic and Indo-European. Linguistic and archaeological considerations*, ed. Christian Carpelan, Asko Parpola, & Petteri Koskikallio, 367–383. Helsinki: Suomalais-Ugrilainen Seura.
- Parpola, Asko. 2013. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology. Revised and integrated 'total' correlations. In *A linguistic map of prehistoric Northern Europe*, ed. Riho Grünthal & Petri Kallio, 119–184. Helsinki: Suomalais-Ugrilainen Seura.
- Peyrot, Michaël. 2019. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European linguistics* 7: 72–121.
- Pinault, Georges-Jean. 2008. *Chrestomathie tokharienne*. Leuven & Paris: Peeters.
- Pystynen, Juho. 2014. Close vowel reduction in Samoyedic. URL: <https://protouralic.wordpress.com/2014/09/22/close-vowel-reduction-in-samoyedic/>.
- Pystynen, Juho. 2017. ü > i, ü in Samoyedic. URL: <https://protouralic.wordpress.com/2017/10/07/u-i-u-in-samoyedic/>.
- Ringe, Don. 1991. Evidence for the position of Tocharian in the Indo-European family? *Die Sprache* 34: 59–123.
- Ringe, Don. 1996. *On the chronology of sound changes in Tocharian*. New Haven: American Oriental Society.

- Salminen, Tapani. 1998. *A morphological dictionary of Tundra Nenets*. Helsinki: Suomalais-Ugrilainen Seura.
- Salminen, Tapani. 2012. Traces of Proto-Samoyed vowel contrasts in Nenets. In *Per Urales ad Orientem: Iter polyphonicum multilingue. Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*, ed. Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi, & Erika Sandman, 339–358. Helsinki: Suomalais-Ugrilainen Seura.
- Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In *The Uralic languages. Description, history and foreign influences*, ed. Denis Sinor, 478–554. Leiden: Brill.
- Schmidt, Karl Horst. 1990. The postulated Pre-Indo-European substrates in Insular Celtic and Tocharian. In *When worlds collide. The Indo-Europeans and the Pre-Indo-Europeans*, ed. T.L. Markey & A.C. Greppin, 179–202. Ann Arbor: Karoma.
- SW = Janhunen, Juha. 1977. *Samojedischer Wortschatz. Gemeinsamojedische Etymologien*. Helsinki: Suomalais-ugrilainen Seura.
- UEW = Rédei, Karoly. 1988–1991. *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.
- Zhang, Fang, Chao Ning, Ashley Scott et al. 2021. The genomic origins of the Bronze Age Tarim Basin mummies. *Nature* 599: 256–261.
- Zhivlov, Mikhail (Михаил Живлов). 2007. К вопросу о реконструкции обско-угорского вокализма. In *Aspekty komparativistiki* 2, ed. A.V. Dybo et al., 281–309. Moskva: Rossijskij Gosudarstvennyj Gumanitarnyj Universitet.
- Zhivlov, Mikhail (Михаил Живлов). 2014. Studies in Uralic vocalism III. *Journal of language relationship* / Вопросы языкового родства 12: 113–148.
- Zhivlov, Mikhail (Михаил Живлов). 2018. Prasamodijskij *s—fonologija i fonetika. URL: https://www.academia.edu/38048065/Прасамодийский_s_фонология_и_фонетика.