CHAPTER NINE

Tocharian Trekkers

Of all the major well-researched Indo-European language stocks, Tocharian is the most intriguing. The two Tocharian languages are among the most recent of the Indo-European languages to have been discovered; they were used along the Silk Road, the easternmost of all the Indo-Europeans; their texts, largely translations, reveal excruciatingly little about their speakers; they are the only language stock to be found exclusively in East Central Asia; and, according to a sizeable body of linguistic opinion, Tocharian's relationship with the rest of its language family makes its location on the extreme east of the Indo-European world totally incongruous. For most of this century we have been able to look at the texts and gaze into the faces of the 'knights with long swords' in Buddhist shrines but still have not the faintest idea who the Tocharians were, what their culture was like, or how they came to be here in the oases of the Tarim Basin. We had the remains of disembodied languages but not their speakers. In this chapter we will present briefly the nature of the Tocharian languages, the basis of our knowledge about them, their position within the Indo-European language family, whatever clues they might themselves offer as to their origins, and why they are so crucial to our understanding of the prehistoric mummies of East Central Asia.

The Discovery of Tocharian

The first Tocharian document to be published was a page containing lines of a hymn to the Buddha which was extracted from a collection of manuscripts held by the Russian consul at Qāshāq. It appeared in 1892 although it would not be read until years later. One of the most recent discoveries of the language was in 1974 when farmers accidentally came across 44 fragments of Tocharian documents at a Buddhist site near Qaratshāhār (Yanqū). Altogether the number of Tocharian documents known runs to 3,640 pages or fragments of Tocharian texts of which the majority are held in Paris (about 2,000), then Berlin (1,170) and the rest in London, St Petersburg, Kyoto, Urumqi and Beijing. To these may be added about 70 additional inscriptions found on walls or other objects (graffiti). We should not be overly impressed with the number of documents since the great majority are fragmentary, possessing less than a single full line of text. The number of Tocharian words known is in the order of 4,500–5,000. This too is less impressive than it might seem since the figure comprises the vocabulary of two languages – Tocharian A and Tocharian B – both of which have borrowed heavily from Indic for their religious terms; essentially, the size of our vocabulary for each Tocharian language is not much larger than that possessed by a 7- or 8-year-old child (although the choice of vocabulary is radically different). Nevertheless, there was at least evidence to indicate that we were dealing with Indo-European languages.

The contents of the manuscripts are overwhelmingly Buddhist and largely translations of texts from Sanskrit or other languages employed in Buddhism. The texts detail the rules of monastic existence, recite the doctrine and law of Buddhism, relate the life of the Buddha in straight narrative form and in translations of Indian plays, and present poetry or even magic spells. In some cases we have bilingual texts in Sanskrit and Tocharian. It is worth emphasizing that the context of discovery of most of our texts is shrines and monasteries, which helps to account for the

<table>
<thead>
<tr>
<th>English</th>
<th>Tocharian B</th>
<th>Old Irish</th>
<th>Latin</th>
<th>Old English</th>
<th>Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>father</td>
<td>pācer</td>
<td>aithair</td>
<td>pater</td>
<td>father</td>
<td>pater</td>
</tr>
<tr>
<td>mother</td>
<td>māicer</td>
<td>māithair</td>
<td>māter</td>
<td>māder</td>
<td>māder</td>
</tr>
<tr>
<td>brother</td>
<td>proicer</td>
<td>brāthair</td>
<td>frater</td>
<td>frater</td>
<td>brāthair</td>
</tr>
<tr>
<td>sister</td>
<td>ser</td>
<td>sūr</td>
<td>soror</td>
<td>soror</td>
<td>soror</td>
</tr>
<tr>
<td>sheep</td>
<td>lāw</td>
<td>ol</td>
<td>ovis</td>
<td>ovis</td>
<td>ovis</td>
</tr>
<tr>
<td>cow</td>
<td>béiw</td>
<td>bo</td>
<td>bōs</td>
<td>cōr</td>
<td>bo</td>
</tr>
<tr>
<td>horse</td>
<td>yukwe</td>
<td>eōh</td>
<td>equus</td>
<td>eōh</td>
<td>hippas</td>
</tr>
</tbody>
</table>

Tocharian greatly simplified the Indo-European sounds, particularly the consonants. Note, for example, that while the Tocharian words for 'father' and 'brother' both begin with the same sound (p), the other Indo-European languages indicate two different sounds. In simple terms, Tocharian did not distinguish between a p and a b nor between a k and a g.

Sanskrit  | yathā bhy aṣṭām duṣchannam aṣṭām samāsthā ṣhadasti
         | evam by abhārīham ottam rāgaḥ samāsthā ṣhadasti

Tocharian | mākte oṣṭā pakwārem aṣṭām swese oṣṭāpe kassūm
           | muni tā mā yātroṣ pāšān ho kassūm enkī oṣṭāpe
           |

as the rain lashes over a poorly-roofed house
so passion lashes over an unpurified spirit.
overwhelming religious bias. Try to imagine what our knowledge of the Romans would be like if we were furnished solely with fragmentary excerpts from the Christian liturgy and had no idea of the existence of Cæsar, Cicero, Virgil or any other classical author. Although we find Tocharian employed in certain secular contexts, e.g. caravan passes, some suggest that the secular language of greatest currency in the north Tarim was the same as that found over a broad area of the south, i.e. a form of Prakrit.

We have given enough public lectures about the Tarim mummies to know that among the most frequently asked questions are ‘Who were the Tocharians?’ and ‘What were the Tocharians like?’ How can one answer that? One way is to refer the reader back to Chapter Two and our account of the oasis state of Kucha as this was one of the main centers of the Tocharians. But this account was based exclusively on Chinese documents that did not recognize the existence of a Tocharian language or people, nor did it isolate specific ethnic identities in the multicultural Tocharian-speaking oases. The evidence of language is our last, albeit imperfect, route to answering the question: ‘Who were the Tocharians?’

How did linguists first crack the code and decipher these previously unknown languages? The original decipherment of Tocharian texts depended essentially on the fact that Tocharian employed a script (Brahmi) that was well known from other languages and that the Tocharian texts themselves were almost exclusively translations of documents that were known in some other language (Sanskrit, Prakrit, Khotanese Saka, Chinese, Tibetan); these provided a ready guide to their contents and a key to the meanings of the words. Occasionally, we find Tocharian translations of Buddhist texts which are not known in the original language of their composition. Manichean is the only other religion to be represented in a single bilingual text in Tocharian and Turkish that dates from the 10th century and was recovered at Turfan. There are also a number of other genres such as grammar, astronomy, medicine and history. Purely secular ‘literature’ has so far been found in a single text, a touching love poem (see opposite). Finally, we have administrative texts concerning the running of monasteries, legal affairs and, as mentioned above, caravan passes. These are extremely useful in that they may also carry dates; the one absolutely securely dated Tocharian text comes from AD 642. The earliest of our texts would date to the 6th, possibly as early as the 5th century AD when the statelets of the Tarim Basin could act independently. They continued into the period of the Tang dynasty (AD 618–907) when the Chinese first regained the Tarim Basin and then lost it in the face of Tibetan and, subsequently, Turkish incursions.

Milestones of the ‘death’ of Tocharian come in a variety of forms. The last mention of the ruling house of Bo ‘White’, which may serve as a rough designation for the ruling Kuchean-speaking aristocracy, occurs in Chinese sources for the year AD 787 and the Uighurs were already predominant in a number of northern oasis towns by 800. Chinese sources are largely silent.

Tocharian Love Poem

[mai] ni cili sos / somo sile / humsime / [?] one taka, mala ra postum / cisi biro masket-ti.

cisi barmume / cisi arti / pelce kaltairuta / solumma sa / mala [e] sile siul waarnai.

tansu pilskanoy / [sanai sarumma / khyau kurtse[!]] latla

waarnai / mai tekoonaum mai管理局.

yamrowrute se / cai ra pi skutai / sarai.

tusa arti erase / cisi anai si fikite.

sotai ci buke / taysra nii wertke / kyanlal-nii-pake po laktetas /
cisi tsiurum samjzise-[u].

There has never been any person dearer to me than you and there never will be any desire.

The love you give me, the delight in you is breath together with life.

This should not change for life.

Thus I used to think: with a single beloved will I live well lifelong without deceit, without pretense.

The god Karman alone knew this my thought.

Therefore, he caused dissolution and tore me from the heart that belongs to you.

He led you away, separated me and had me partake of all sorrows.

The joy I had in you he took away from me.

Portion of a Tocharian caravan pass (literal translation by D. Q. Adams)
siklit pilkum yasino se

[the mountain commander writes to Yusi: thou my command—by thus do that [the] Qashqarian Bludhdamitra outside goes: him with twenty: asses three: horse one: this outside permit: than this more not permit ten[th] year: regular-period-off. fourth montba

on the Tarim Basin for the period of the 9th and 10th centuries and by the 11th century we find at least the northern oases, especially the Turpan Basin, clearly under Uighur rule. We have sufficient evidence of Tocharian texts with either glosses in Turkish or with indications that their copying had been commissioned by Turkish speakers that we can imagine that Turkish spread first as a vernacular, replacing the languages of the earlier Tocharians, until Tocharian survived only as a liturgical language which then itself gradually disappeared as the population abandoned it in favour of Uighur; eventually, the population also adopted the Islamic religion.
The Tocharian Languages

We have seen that written documents reveal the existence of two Tocharian languages, commonly designated Tocharian A and Tocharian B. The texts themselves are found across the northern route of the Tarim Basin from Maralbeshi (Bachu) on the west as far east as Beizilik on the east which helps us define the known limits of the Tocharian world. The specific sites of Tocharian A documents are confined to the east where they are found at the religious sanctuaries of Sharchoq near Qarashahar, the ancient kingdom of Agtu, and hence Tocharian A is also known as Agtene or East Tocharian. Documents in Tocharian A are also found farther east in Buddhist sites in the vicinity of Turpan such as Qarakhoja and Khoko, and at Tuyuq, Sangim and Beizilik, all excavated as part of the German expeditions at the beginning of this century. Tocharian B documents are found across the entire region. They have been recovered from Maralbeshi, in various religious establishments near Kucha, i.e. Duldur Aqur, Quntura, Subeshi and Qizil. It is for this reason that Tocharian B is also known as Kucheite or West Tocharian, but the latter is something of a misnomer since Tocharian B documents are also known across the east, generally in the same sites as Tocharian A, i.e. Qarashahar, Turpan, Khoko, Tuyuq, Sangim and exclusively at Murtoq.

How great is the difference between Tocharian A and Tocharian B? Substantial enough in vocabulary and grammar that it is doubtful that they were mutually intelligible languages at the time they were committed to writing; although one can always find identical or extremely similar words between the two. Donald Ringe suggested that the two Tocharian languages were about as similar as Old English and Old High German (D. Q. Adams prefers the contrast between Italian and Romanian). Let us give both of these comparisons a brief run by first counting to ten and then listing some basic kinship words in the two Tocharian languages and the languages cited by Ringe and Adams (the Tocharian numerals can be given both in a masculine form and a feminine form, but only the masculine will be given here).

<table>
<thead>
<tr>
<th>English</th>
<th>Tochar A</th>
<th>Tochar B</th>
<th>Old Eng</th>
<th>Old High German</th>
<th>Italian</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>sus</td>
<td>se</td>
<td>an</td>
<td>ein</td>
<td>uno</td>
<td>uno</td>
</tr>
<tr>
<td>two</td>
<td>se</td>
<td>wi</td>
<td>twa</td>
<td>zwei</td>
<td>due</td>
<td>dot</td>
</tr>
<tr>
<td>three</td>
<td>tre</td>
<td>mai</td>
<td>fria</td>
<td>drei</td>
<td>tre</td>
<td>trei</td>
</tr>
<tr>
<td>four</td>
<td>siwar</td>
<td>siwer</td>
<td>fener</td>
<td>quattro</td>
<td>patra</td>
<td>patra</td>
</tr>
<tr>
<td>five</td>
<td>pani</td>
<td>pisi</td>
<td>fumf</td>
<td>conque</td>
<td>cinci</td>
<td>cinci</td>
</tr>
<tr>
<td>six</td>
<td>sàk</td>
<td>skas</td>
<td>siex</td>
<td>set</td>
<td>sìse</td>
<td>sete</td>
</tr>
<tr>
<td>seven</td>
<td>spit</td>
<td>sakt</td>
<td>seoson</td>
<td>sette</td>
<td>sàpte</td>
<td>sete</td>
</tr>
<tr>
<td>eight</td>
<td>okar</td>
<td>okt</td>
<td>ahto</td>
<td>otto</td>
<td>opt</td>
<td>opt</td>
</tr>
<tr>
<td>nine</td>
<td>niu</td>
<td>niu</td>
<td>nigon</td>
<td>nòve</td>
<td>novã</td>
<td>novã</td>
</tr>
<tr>
<td>ten</td>
<td>sàk</td>
<td>sàk</td>
<td>tiêns</td>
<td>diêci</td>
<td>zecê</td>
<td>zecê</td>
</tr>
<tr>
<td>father</td>
<td>pàcar</td>
<td>pàcar</td>
<td>fader</td>
<td>padre</td>
<td>tata</td>
<td>tata</td>
</tr>
<tr>
<td>mother</td>
<td>mòkar</td>
<td>mòkar</td>
<td>mòtór</td>
<td>mòder</td>
<td>marô</td>
<td>marô</td>
</tr>
<tr>
<td>brother</td>
<td>pracar</td>
<td>pracer</td>
<td>brodór</td>
<td>fraitel</td>
<td>friàle</td>
<td>friâle</td>
</tr>
<tr>
<td>sister</td>
<td>par</td>
<td>ser</td>
<td>sweiotr</td>
<td>sorîla</td>
<td>sorîlã</td>
<td>sorîlã</td>
</tr>
</tbody>
</table>

In most of these cases, the words in the different languages are cognate with each other, i.e. they come from the same ultimate Proto-Indo-European origin (Romanian employs the children's forms for 'father' and
mother' which could be paralleled in the other languages). On the other hand, there are also differences in vocabulary, at least where we have the texts for the words, between the two Tocharian languages, even in some fairly 'basic' vocabulary. Here are a few examples with the comparative words alongside for good measure (foot of previous page).

And when we come to the grammar of the two languages, here too we find marked differences. We can take the word for 'horse' (c.f. Latin equus) as an example.

<table>
<thead>
<tr>
<th>Case</th>
<th>TocharA</th>
<th>TocharB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>yak</td>
<td>yakowe</td>
</tr>
<tr>
<td>Oblique</td>
<td>yøk</td>
<td>yøkwe</td>
</tr>
<tr>
<td>Instrumental</td>
<td>yøkoyo</td>
<td>yøkwe-a</td>
</tr>
<tr>
<td>Pedalive</td>
<td>yøkt</td>
<td>yøkwe-a</td>
</tr>
<tr>
<td>Genitive</td>
<td>yøk</td>
<td>yøkwen-te</td>
</tr>
</tbody>
</table>

In short, we are not dealing with two dialects that closely resemble each other, but something more on the order of two different languages. This is an important point because it suggests that these languages had been separated from one another for a considerable period, estimated by linguists as possibly somewhere between 300 years and a millennium, before our earliest texts. This would indicate a lower date for the break-up of Proto-Tocharian, the language ancestral to Tocharian A and B, at very roughly 500 BC.

A second major difference between the two languages is their geographical situation. We find Tocharian B texts across the entire northern rim of the Tarim Basin (about 1,100 km or 700 miles) while Tocharian A seems to have been confined to the area between Karasahari and Bezaklik, a linear distance of only 300 km (200 miles). In the case of Tocharian B, regional differences have suggested the existence of three dialects: the first (western) in Kucha and the area surrounding it, the central dialect of the Karasahar region and, finally, the eastern dialect of the Turpan oasis in the east. Werner Winter has proposed that the central dialect, best represented by texts from Shorchuq, was the prestige dialect of the monastery of Yurpokska which eclipsed Kucha after it fell in the mid-7th century.

The third, and one of the most critical distinctions between the Tocharian languages, is the context in which they were used. While both languages were employed in the translation of Buddhist documents, only Tocharian B occurs in secular and administrative contexts as well. This has led to the well-supported supposition that what we have of Tocharian A is the remains of a liturgical language, probably a dead liturgical language at that, while Tocharian B was the only living language in the period from which we draw our texts. We find instances, for example, where a monk, copying out a Tocharian A text, would come across words that would be unfamiliar to him, and feel the need to write their definitions in his own language (Tocharian B). This was a widespread practice in the Middle Ages and we could just as well point to instances where Irish monks, confronted with a section of tricky Latin text, scribbled the meanings in Old Irish in the margins of their manuscripts or immediately above the difficult words. We also find accounts of the history of Kucha, the centre of the western language and also of Buddhist learning in the Tarim region, recorded in Tocharian B as far east as Murtoq, in an area where we otherwise find Tocharian A texts. And at Sanging, again in the east, we discover a text instructing the novice in Sanskrit where the words are glossed in Tocharian B.

One of the ironies of the Tocharian languages is that the living and the dead language do not behave as expected: Tocharian B, the 'living' language, was on the whole the more conservative of the two languages. Although both Tocharian languages have seen fairly brusque losses of their earlier endings, Tocharian A reveals far greater loss of syllables and endings, e.g. Tocharian A yøk 'horse', Kopti 'hundred', and ksm 'wagon' but Tocharian B yøkwe, kurt, and koko. In crude terms, there is unexpectedly more left over of the Proto-Tocharian forms in the 'living' Tocharian B than in 'dead' Tocharian A.

We have then two languages, considerably different from one another, one of which may have been purely a liturgical language. Explaining their geographical context becomes all the more complicated when one considers that the distance between Kucha and Karasahar is only about 300 km (200 miles). How could two languages which were geographically so close have diverged so much from one another? Werner Winter has argued that Tocharian A did not actually develop in the region in which we find it used, but that it was merely the language of the Buddhist mission to the Turks. The Tocharian A manuscripts were prepared under the auspices of the Turks who settled in the eastern part of the Tarim Basin in territories that were occupied by Tocharian B speakers. This would suggest that Tocharian A, when it was at home, was situated somewhere between the Tocharian B speakers of the north Tarim and the early Turks to their north and east. This, of course, would imply that the Tocharian languages were once far more widespread than their surviving manuscripts reveal. This is no surprise and one can imagine that the A and B languages are merely the two surviving languages of a whole chain in which the intermediate languages had died out. This hypothesis will become useful when we are searching for the linguistic identity of some of the mummies.

Tocharian C?

While Tocharian documents are well attested in the northern part of the Tarim Basin, no Tocharian documents have been discovered along the southern Silk Road. Here, as we have already seen, we find documents largely in Prakrit, a Middle Indic language, that served both as a liturgical
and administrative language, and in Khotanese Saka, an Iranian language. A sizeable group of documents was uncovered in the ancient territory of Krorian (Kroraina, Loulan, later Shanstan). These were among the 'treasures' recovered from the highly productive site of Niyâ, excavated by Aurel Stein, which lay on the borders of Khorasan, while other documents were discovered at Andir and in the vicinity of Krorian itself. The Prakrit documents, however, revealed various strata of vocabulary. Although written in a language of northwestern India, their authors had frequent occasion to use words in a different language. About 40 of these 'foreign' words are borrowed from Persian and many of the proper names are Indic but there remain about an additional 100 words and 1,000 personal and place names that cannot be explained as either Indic or Iranian. These exotic words fill out many semantic categories ranging from agriculture, land use, textiles, material culture, transportation and social organization. That the words fill so many semantic categories and that they also include so many local, place and personal names suggests that these words derive from the 'natives' of Krorian and, therefore, we can refer to this segment of the vocabulary as Kroranian.

What was Kroranian? Thomas Burrow examined these peculiar words in the 1930s and came to a conclusion that still finds general support today – that the native vocabulary is best explained as some form of Tocharian. One has to be vague here since the linguists are really in difficulty when it comes to the Kroranian vocabulary. The problem is that the Prakrit documents (and Kroranian words) are administrative texts but we simply do not find corresponding texts that deal with the same things in Tocharian A or B. It is something like being given a handful of pages ripped from an automobile manual detailing the repair of a carburetor and told to use it as a key to translating Shakespeare's sonnets. Unsurprisingly then, only a handful of words have been compared so far between Kroranian and the two Tocharian languages. The local Kroranian word for 'district' kilme is certainly to be compared with Tocharian $\bar{A}dyme$ 'district'. Much less certain are some of the other words proposed. Referring to a camel, the Prakrit text describes one as amkiata, which Burrow compared with the Tocharian (A and B) word aknista 'ignorant', i.e. an untrained camel. Among the names of agricultural products, the author had to resort to a local word, ogana, which was compared with Tocharian B oko 'fruit'. Less plausibly, maka 'some commodity that could be sent a long distance as a tax' was set beside Tocharian maka 'milk' (although one would be a bit surprised to see such generic terms turning to the local patois; nor can one readily imagine a long-distance milk shipment in the Tarim Basin).

Far more convincing is the fact that the exotic words tend to reflect the Tocharian phonetic system and contain what would appear to be Tocharian suffixes. For example, in Tocharian $\bar{A}$, one of the ways of forming an adjectival is to add the suffix {-eci, eci} (e.g. etal 'man' but etlisci 'belonging to the man, male'). In the Prakrit documents we find a similar suffix {-meci as in kilme 'district' but kilmeci/mecici 'belonging to the district' or Krorami$\dot{e}$ ci 'Kroranian'). In short, it is as if one stumbled across a document in English which read that 'Brothers Schwabe and Schuwer behaving dicht and schwienisch during their morning prayers'. The two proper names and one of the 'strange' adverbs all begin with a sch, which is not admitted in English (except in German loanwords), while the other adverbs, even if we didn't know their meaning ('apish' and 'swinish'), have the same non-English -isch suffix which, were we to know German, we would be able to identify easily enough.

Burrow concluded that in the southern Tarim there existed traces of a Tocharian language. Although the hard evidence was meagre enough, Burrow also suggested that it reflected a different dialect than that found in the northern Tarim Basin (we would need to know much more about the English language before we could pronounce definitively on this), which we might call Kroranian or Tocharian C. What is most interesting is that the Prakrit documents date from the 3rd century AD i.e. they are about 300 years older than any of our other Tocharian documents. That they are replete with what are taken to be Tocharian personal and place names and what are apparently local names for agricultural commodities suggests that they reflect the language of the natives of the region who had converted to Buddhism as it spread along the southern route through East Central Asia. This places Tocharian C in the region earlier than the Buddhist mission to the Tarim Basin. That the documents contain so few Iranian loanwords (which, as Burrow suggests, are more likely to have been picked up en route from India to the Tarim Basin) indicates that Tocharians may have also preceded Iranians in their occupation of the central and western regions of the southern Tarim Basin. It might be added that analysis of loanwords specifically associated with Buddhism in the other two Tocharian languages (Tocharian A and B) suggests that the Saka were the earliest agents for the spread of Buddhism here. The Tocharians, in short, make a case for a people or, at least, a language that preceded the Buddhist mission, the Indo-Aryans and, more importantly, probably the Iranians into the southeast of the Tarim Basin. And of still greater interest, it places the earliest evidence for Tocharian in the same region where we find the earliest evidence for Caucasoid settlement of East Central Asia as well as many of our earliest and best-preserved mummies.
By Any Other Name?

What did the people who left the Tocharian documents call themselves? On the one hand, this is a fairly easy question: on the other, it presents one of the goriest academic battlefields imaginable and, so as not to sicken the general reader, we have set much of our discussion in an extended note at the end of this book where the braver members of our audience, those who sit at the back of lecture halls and torment speakers with questions about the jumble of names that constitutes the 'Tocharian problems', may sate their lust. Let us start with what we know.

Tocharian B documents make it explicit that at least some of the people who wrote them regarded themselves as Kuchans. The name of the state was Kosi or 'Kusi' which is rendered today in Mandarin as Qosi (in the early medieval period this would have been pronounced *kau-*). The adjectival was kusînd and we even have a document with the phrase kusînd orooceri tênta yaitkorsa 'by the order of the great king of Kucha'; the same appellation is provided in a Sanskrit document which refers to the Kusimahârata, 'great king of Kucha'.

As we have seen above, the region of Karashahr where Tocharian B texts have been recovered was known in Sanskrit documents as Agni and in Chinese as Yanqi (Middle Chinese *tsan-gi* or *tsan-gi*). The name is also well represented in other languages, e.g. Khotanese Argi and Buddhîs Hybrid Sanskrit Agnideia ('Agni-country'). Some have maintained that both names were derived from words meaning 'white' or 'silver', i.e. both Kuchans and Agnans regarded themselves as 'the brilliant ones', but there are sufficient reasons to be wary of such claims and pursuing them gets us no closer to their identities.

Now it would be convenient to leave things as they stand, i.e. designating the western language as Kuchan and the eastern as Agnian, were it not for several other texts which lead us into a labyrinth of linguistic and historical arguments. The attentive reader will by now already seen at least one of the problems: if the speakers of the two languages called themselves Kuchans and Agnans, then why do we call them Tocharians? And hasn't we already mentioned the Tocharians with respect to the Yuechi conquest of Bactria and the establishment of the Kushan empire? Are all these the same people, i.e. were the Yuechi speakers of Tocharian languages who carried their language in Bactria in the 2nd century BC? We will set out the problem so that the reader can understand the ultimate implications of such questions. In order to do this we will have to engage in some wordplay and, until further notice, we will refer to the languages attested in the north Tarim basin as KA (Kuchan-Agnian) rather than Tocharian.

As we have seen, the latest of the KA texts were prepared under the auspices of the Uyghurs whose own Turkic language ultimately replaced that of the earlier inhabitants of the region. About AD 800 an Uyghur translation of a Sanskrit drama was prepared and in the colophon (that part of a manuscript referring to the circumstances of its creation), one reads in Uyghur: 'The sacred book Mattreyasamiti, which was composed by the Bodhisattva gurñ acaña Aravinda, native of Agnidesa, in the togrî language from the Sanskrit language, and which has been translated by gurñ acaña Prajñâraksita, native of Il-baliq [the Uyghur capital of Khocho], from the togrî language into the Turki language'. In a nutshell, the colophon tells us that a Sanskrit drama was first translated from Sanskrit into the togrî language of Agnidesa (and we know that this means the Agnian kingdom) and then from togrî into the Turki language of Il-baliq by two learned gurus. Armed with this equation, in 1908 E. Sieg and F. W. K. Müller reasoned that the apparent Uyghur designation of Agnian, togrî, was very close to the name of the well-known ethnic group Tökhâr (in Greek), Latin Tochari, Sanskrit Tukhâra or Chinese Tüokuans, the name applied in ancient sources to Bactria, i.e. Tocharistan. It has been further suggested that the same name can be found in other ancient sources, e.g. Prolem's gazetteer of the known world, dated to the 2nd century AD, lists a Thaugouroi in Gansu, a Tukhansiai north of the Imaus (Himalayas) and a Tukhansiai in the vicinity of Issyk-kul. Peoples with a name sounding like 'Tocharian' seemed to be everywhere from the borders of Gansu to Bactria. To give an idea of where such observations can lead, consider as examples the following two proposals by A. K. Narain and W. W. Henning.

As we have just mentioned, the people who emerge as Tocharians in Western sources are often equated with a branch of the Yuechi of Chinese sources who were driven first from the Gansu borderlands by the Xiongnu, then farther west by the Wusun, arriving at the Oxus, and going on to conquer Bactria and establish the Kushan empire. Narain argues that once one accepts the equation Tocharian = Yuechi, then one is forced to follow both the Chinese historical sources (which for him would propel the Yuechi back to at least the 7th century BC) and the geographical reference of their first cited historical location (Gansu) to the conclusion that they had lived there 'from times immemorial'. Narain infers that they had been there at least since the Qia culture of e. 2000 BC and probably even earlier in the Yangshao culture of the Neolithic. This would render the Tocharians as virtually native to Gansu (and earlier than the putative spread of the Neolithic to Xinjiang) and Narain goes so far as to argue that the Indo-European themselves originally dispersed from this area westwards. Seldom has a tail so small wagged a dog so large.

Rather than starting from the east, Henning found the origin of the Tocharians to lie in the west. He picked up their historical trail on the western borders of Babylon where, e. 2100 BC, the kingdom of the Akhadians was being harassed by the barbarian Guti 'whose language is queer' and whose neighbours included the Tukhâra. Henning speculated that both of these peoples had probably come from farther north, south Russia to be precise, by way of the Caucasus very much in the same manner that the later Scythians and Sarmatians poured down into the Middle East in the
1st millennium BC. Why they should be associated with the peoples of the Tarim Basin rested on several lines of evidence. First and foremost was the apparent similarity of these names with those found 2,000 years later in the Tarim. At that time, according to Henning, the pronunciation of the Chinese morphosyllables that are now pronounced Yuezhi should have been approximately ‘Gu(t)-t̚i’, which would provide a perfect phonological correspondence. In the KA language, Guti would have been rendered Ku-ti, and hence be equivalent to Kuchean. As for the tagri mentioned in the Uyghur colophon, Henning believed one need look no further than the name of the Tukris who had been neighbors of the Guti in western Persia and hence had given their name to both the tagri of the northern Tarim and the Tocharians of Bactria.

The presumption that there was a connection between two names found in Akkadian documents and names encountered over two millennia later in the Tarim Basin did not elicit widespread support. Henning suggested that the variation in endings found among the names of Gutian kings might be explained by the change in endings which themselves might be associated with those employed in Tocharian documents some two and a half thousand years later, an enterprise of extreme dubiousness. He also made passing reference to similarities between ceramics found in Iran and in the Tarim Basin, although these currently play no part in discussions of the prehistory of the Tarim Basin, and the maintenance of ceramics, especially painted wares, among supposedly nomadic peoples wandering from Persia into the Tarim would be regarded as exceptional indeed on every count. Of greater detriment to such a theory is that Henning accepted a reconstructed Chinese pronunciation of Yuezhi as ‘Gu(t)-t̚i’ when, in fact, it is commonly reconstructed now as ‘ngiuot-t̚i̯g’ which makes it a far less transparent correspondence.

How one interprets these various references to Tocharians and words similar to Tocharians does make an obvious difference to how one interprets the ethnic history of East Central Asia. But any conclusions are so tentative that we will cut to the chase and try to present the main kinds of alternatives conclusions in terms relevant to our search for the identification of the Tarim mummies.

The first type of result we can call the Iranian hypothesis. It begins with the observation that there is no evidence whatsoever that the historical Tocharians, who are associated with the conquest of Bactria, and who included in their number the people who were later to be called the Kushans, ever spoke a language similar to KA; all the evidence of the Kushan empire indicates that those who administered it employed an Iranian language (or Pātrkāt). As for the Yuezhi, there is no firm evidence from any Chinese source as to the language that they spoke. Prior to the Xiongnu political dominance of the Tarim Basin, the region may have been under Yuezhi control but this did not entail any linguistic domination, although it may have been yet another vehicle for Iranian loanwords into the KA languages. The basic urban population, or at least their ruling classes, remained KA speakers irrespective of whether they were ruled by Yuezhi, Xiongnu or Chinese; only the Uyghur, who actually settled in the region and occupied towns themselves, were able to effect the linguistic assimilation of the KA speakers. This hypothesis receives some coincidental support from descriptions of the Greater Yuezhi which indicate that they, like other Iranian-speaking peoples described in ancient literature, permitted women to hold high-status positions. Moreover, they are regarded as the same as the Iranian-speaking Parthians with respect to land, customs, produce and coinage, and it has been suggested that their success in bringing the region under their own power was due to the fact that they shared a similar ethnic perspective and language. While it is not greatly important to us what language was carried into Bactria by the Yuezhi, this hypothesis suggests that the eastern part of Central Asia was dominated by Iranian speakers during the last centuries BC, possibly earlier. This hypothesis receives additional support from the evidence for cultural borrowing during the Bronze Age between Iranians and Chinese which would require very early direct contacts between the two populations.

We can refer to the second hypothesis as the Tocharian gambit. It runs something like this. The Yuezhi, who occupied the area from Gansu west to the Tarim Basin, spoke KA languages. By the time those who migrated west arrived in the Oxus, they had shed KA for an Iranian language or simply took up the expedient course of many foreign rulers and employed the local Iranian language for their administration. From our own particular perspective, this would mean that both the nomads of the eastern border of East Central Asia and the oasis-dwellers of the Tarim and Turfan basins and the Lopnor region spoke a KA language, at least by the first centuries BC (we can imagine here a situation something like the Arabs who may be either settled farmers or nomadic Bedouins). As we also have some Tocharian loanwords in Chinese that date to about the 3rd century BC (possibly much earlier but that cannot be easily demonstrated), we can have a Tocharian rather than an Iranian world adjoining the Han.

No matter how these hypotheses develop, the starting point is the Yuezhi on the borders of Gansu (and it may be that their location 'between Dunhuang and the Qilian Mountains' actually indicated that they were situated farther to the northwest, i.e. between Dunhuang and the Tāngri Tagh). So which, if any, of our mummies belonged to the Yuezhi, whose language we are most uncertain about in the first place? It is time to introduce an admittedly hallowed archaeological spin to the whole affair. The Yuezhi of the Gansu corridor have an historical existence, not an archaeological presence, i.e. there is so far absolutely no evidence for the camps or burials of the more than 400,000 Yuezhi recorded in Chinese texts anywhere that historians have positioned them. While we may not expect much from pastoral nomads, we should remember that the Saka and Wusun are both identified by their burials in the Ilı River region and we also cannot dismiss the Gansu corridor as a terra incognita, as here we find the
settlements and cemeteries of the Siba culture, an Early Bronze Age culture whose physical type would appear to be Mongoloid, in the same territory where we would later expect the Yuehi. The only way we can discuss Yuehi is to extend the name over much broader areas of the Tarim and Turpan basins than the historical records allow. We are engaged in moving names across a map without any evidence of the people. If the Chinese histories had never mentioned the Yuehi on the borders of China, no archaeologist would have had the slightest reason to postulate their existence. The Yuehi are 'ghosts', summoned up by historians to torment archaeologists. We have no need to look for ghosts; we have the remains of real people with at least some evidence of their material culture.

Linguistic Prehistory

Embodied in any language is evidence of its past. The preceding sentence of nine words contains five words which are derived from Germanic (in, any, is, of, its), three which derive from Latin via French (language, evidence, past) and one which combines a Latin prefix with a Germanic root (embodied). That the most basic verbs, articles and prepositions are all Germanic would suggest a relative chronology in which English was originally a Germanic language that was later heavily influenced by a Romance language.

An examination of the Tocharian languages (we can now come up for air and revert to this designation rather than KA) also reveals something of their past although we cannot work back too far here. Curiously enough, the impact of Chinese on Tocharian is not very significant (e.g. Tocharian B căne 'price' derives from the Chinese word for 'money', the (modern) qiān with the medieval pronunciation 'dz'am; also measurements of capacity; and, as D. Q. Adams has recently indicated, the word for 'rice' nor does Turkish provide a significant source of Tocharian vocabulary. On the other hand, Middle Iranian languages current in the Tarim Basin and surrounding areas in the 1st millennium AD provided a major and predictable source of loanwords to the Tocharian languages. From Khotanese Saka, the language of the southern Tarim Basin, we find loanwords such as asana- 'worthy' borrowed as Tocharian B asām 'worthy', while the Sogdians who acted as one of the principal middlemen in the transportation of goods along the Silk Road contributed the word mādā 'wine' (cognate with our English mead) which was taken up as Tocharian B mar 'alcohol'. Somewhat earlier contacts with Iranian languages can be seen in words such as tām, 'arrow', which was apparently borrowed from a northeast Iranian language such as Avestan (zāmān- 'weapon'). Not all borrowings came by way of Iranian since there is also evidence of Indic borrowings (as well as the direct importation of Sanskrit in Tocharian religious literature). These terms refer to aspects of Buddhist life which would have been foreign to the pre-Buddhist Tocharians, e.g. Tocharian B a stanza 'nun', aṣāra 'teacher' and saṃāne 'monk' which all find their origins in words from Gândhari (the language of Asoka's inscriptions) of the 2nd century BC to the 4th century AD.

Tocharian specialists are generally reluctant to admit that there was any flow of Indo-Iranian vocabulary into Tocharian earlier than the 1st millennium BC, a time when words such as Tocharian A pọrati, Tocharian B pọrati 'axe' may have been borrowed from some Iranian language (we find Ossetic firat 'axe'). Tocharian also employs a word for 'iron' (A aṅcu, B ecucu) which some compare with similar words such as Ossetic andan 'steel'. While such evidence can only be employed positively, i.e. it cannot demonstrate that contacts did not exist, there have been few to suggest that the ancestors of the Tocharians were in close contact with their Indo-Iranian neighbours before c. 1000–500 BC. To go back any earlier takes us into one of the most disputed areas of Tocharian studies.

Real 'Europeans'?

The relationship between Tocharian and the other Indo-European stocks or subfamilies and the issue of how 'European' the Tocharian languages actually are has dogged Tocharian studies since its inception. If we reflect again on the various Indo-European language stocks, we see that that which embraces the Indic (and Nuristani) and Iranian languages is the most certain higher-order group in Indo-European. By this we mean that we can reconstruct a linguistic stage between the development of Iranian and Indo-Aryan on the one hand, and Proto-Indo-European on the other, namely, Proto-Indo-Iranian. In short, we get from Proto-Indo-European to Sanskrit or Avestian only by way of Proto-Indo-Iranian, and any discussion of Indo-Aryan or Iranian origins will include the geographical staging area of Proto-Indo-Iranian. The majority of, although not all, Indo-Europeanists would also accept a Balto-Slavic stage before the independent development of the Baltic and Slavic languages. Now both these subgroups of Indo-European are satem languages (i.e. they changed a 'hard' g or k to j, z or s) and the existence of other unique linguistic features shared between these satem languages suggests some degree of proximity between the two groups, i.e. in some way there was a continuum of Balto-Slavic-Indo-Iranian which experienced a number of common innovations not experienced by the other Indo-European stocks. Tocharian, as we have seen in Chapter Three, belongs to the centum languages which are defined by their hard pronunciation of what we reconstruct as palatal velars (t'g, t'k, etc.). It is generally, although not universally, presumed that the centum languages did not participate in the innovations experienced by the satem languages; being a centum language, however, does not tell you any more about its prehistoric relations. For that, linguists have sought other criteria among the various centum languages to determine their correct positions on the Indo-European family tree. And it is a truly amazing array of models that they have proposed. We list their putative relationships below on a west to east axis.
1. Tocharian is most closely related to Italo-Celtic, i.e. its closest linguistic relatives are to be found in the far west of Europe. There were those who claimed that it was specifically closest to Celtic and one scholar even went so far as to call the ‘other’ Celtic languages (which, despite the tartans, it emphatically is not). Geographically, this theory places the pre-Tocharians somewhere on the eastern fringes of the early Celtic world, i.e. central Europe.

2. Tocharian is most closely related to Germanic. Geographically, this normally sets the pre-Tocharians somewhere to the south and east of the early Germanic languages of northern Europe.

3. Tocharian is most closely related to Italo-Celtic, Germanic and Balto-Slavic, i.e. it was one of the languages of the so-called ‘Northwest European group’ and this theory places the pre-Tocharians somewhere on the southeastern margin of these other languages.

4. Tocharian is related most closely to Balto-Slavic and it falls somewhere between Balto-Slavic and Greek, i.e. it was a language whose closest relations are to be sought in central and eastern Europe.

5. Tocharian is most closely related to Greek (or established later relationships with Greek after earlier relationships with languages farther to the northwest). This relationship need not be geographically translated into a Greek origin since many would presume that pre-Greeks entered Greece from the north; a southeast European origin should be sought in this case.

6. Tocharian is most closely related to Thracian (a Balkan language) and Phrygian (employed in Anatolia). Here again, a southeast European origin, possibly one that found the pre-Tocharians on the northwest corner of the Black Sea, may be envisaged. It might be added that both these languages are so poorly known that this hypothesis is barely discussable.

7. Tocharian belongs to the Indo-European peripheral languages which comprise Celtic, Italic, Phrygian, Anatolian and Tocharian, i.e. the expansion of the Indo-European languages was centrifugal from a centre that found the Tocharians on the outer eastern rim of language expansion. Here, Tocharian’s relationship with Italic and Celtic, for example, does not have geographical connotations other than that they all were on the outer rim of the Indo-European world.

8. Tocharian is a relatively independent language stock not closely related to any other stock. Generally, such a conclusion presumes that Tocharian departed early (after Anatolian) from the continuum of late Indo-European dialects, thereby it shares few if any innovations with other Indo-European stocks. Such a model is geographically non-specific provided that at an early date the pre-Tocharians take themselves away from the other Indo-European stocks.

The Measure of a Language

To some degree, all these theories for Tocharian’s relations are still very much alive and kicking, so much so that hardly a subsequent sentence in this chapter could not be vigorously challenged by a competent Tocharian specialist or Indo-Europeanist. How can this be? The answer lies partly in the nature of the evidence and partly in the methods by which linguists seek to establish the proximity of one genetically related stock with another. The hard linguistic ‘facts’ are the comparative descriptions of the phonology (sounds), morphology (grammar) and vocabulary of Tocharian with other Indo-European languages. Two languages, genetically related, may share a common feature because 1) they both inherited it from the proto-language (even if all the other languages have lost it); 2) they have both created the feature independently in their own language; 3) the immediate linguistic ancestor of both created the feature at some time after the dissolution of the proto-language; and 4) one language innovated and passed the new feature onto a geographically contiguous language. We have, in effect, already investigated situation 4 where we have found late loans between the Indo-Iranian languages and Tocharian. We now need to look briefly at the other situations.

Let us propose that Tocharian and Celtic are especially closely related as a ‘screw dog’ to see how these different principles work. We can see that a similarity between two stocks that is determined by situation 1 is of no geographical utility because the two languages are, like all the other members of the same language family, only related by way of the proto-language itself. That Tocharian A yak ‘horse’ is cognate with Old Irish eoch
'horse' tells us nothing of a common geographical origin since cognates for the same word for 'horse' can be found in almost all other Indo-European languages (e.g. Old English eoh, Latin equus, Greek hippoc, Sanskrit ārasing); it only shows that they all derive from a common proto-form, not that there is any special relationship between, say, Irish and Tocharian.

This is easily understood, but how do we distinguish between this situation and that described in situation 3 where the two languages in question were the only languages that shared a particular set of cognate words or grammatical forms? This is one of the grounds on which linguists have sought to demonstrate that Tocharian has a more exclusive or unique relationship with some rather than other Indo-European languages. Some of the evidence of vocabulary shared exclusively between Tocharian and other languages has been assembled by both A. J. Van Windekens and Douglas Adams.

What do these lists convey? They at least illustrate the reason why we may find some linguists attempting to locate the ancestors of the Tocharians somewhere, say, between the Germans, the Balts and the Greeks (and some of the others such as Celtic look far less impressive), but do they offer sufficient grounds for any conclusion? Even Douglas Adams admits that one cannot tweak these figures into what might pass for statistical probability, and they seem to throw up as many problems as they try to resolve. For example, why should Tocharian share about half as many isoglosses with Iranian as Indo-Aryan when the shared vocabulary should have occurred before the collapse of Proto-Indo-Iranian? The probable answer is that we have preserved for us a vastly larger Indo-Aryan vocabulary than we have for early Iranian and hence a far better chance of finding cognates. The same would go for Greek and Germanic, the first with a truly enormous early vocabulary, and the second with an impressive number of well-attested and well-studied languages. The only rule evident here is that the greater the overall preserved vocabulary, the greater the chance of finding cognate 'hits' between Tocharian and some other language. Douglas Adams has attempted to rectify this situation by recomputing the figures as if each language were an equal contributor to our reconstructed Proto-Indo-European vocabulary (his 'revised' computations above). But the results will almost invariably be contradictory. For example, there is a unique Tocharian-Germanic isogloss in that 'Tocharian A hokon 'boat' and Tocharian B ollom 'boat' would appear to be cognate with Old High German skalma 'boat'.

Fine, we move the pre-Tocharians somewhere towards northern Europe. On the other hand, when we look at the outcome for the cognate sets for the word for 'head' (drod), we find that Tocharian A (drod 'head'), B (adiim 'head') aligns itself with Baltic (Lithuanian arūdà), and Indo-Iranian (Avestan aro-, Sanskrit arā-) and does not participate in the prefixing of this word with a d-sound found in Celtic (Old Irish dár), Italic (Old Latin decuma), Germanic (Old English dēr), and Greek (dēkē). Here it is peripheral perhaps and certainly nowhere near the early Germans.

It may be argued that comparison of grammatical and phonological traits should be of greater value in that we generally have enough of any major language stock to evaluate these features on an equal basis. Douglas Adams incorporated the Tocharian data into an earlier study that utilized 30 phonological and 44 morphological traits to examine the similarities among the various Indo-European dialects. His data revealed that Tocharian shares features in order of frequency with Greek, Armenian, Germanic, Italic, Balto-Slavic and Celtic (while showing negative correlations with Indo-Iranian). But these similarities embrace a wide variety of types: many may be shared retentions of features rather than common innovations and, what we take as innovations may be independent as well as shared. It is, therefore, the nature of the evidence rather than the quantity that counts in establishing a relationship between two stocks, and unfortunately there is simply too little consensus here to make much of a case for any special geographical relationship between Tocharian and another Indo-European language stock.

As anyone familiar with clustering techniques is aware, just about any body of data will permit itself to be clustered no matter how meaningless it is. We could prepare a tree-diagram that would show how Beethoven was 'closer' to Haydn than Bach; adding Hoagy Carmichael, Scott Joplin and Elton John to the same tree is always possible but what would the results mean? Does it matter anyway? It certainly does and the 'saudiced archaeologist' gambit is not going to get us out of this mess.

The Flight of the Arrow

The most powerful weapon in the arsenal of anyone attempting to establish the origins and dispersion of an ethnic or linguistic group is a felt-tipped pen, preferably red (but a black one will do at a pinch). Consider for a moment...
how the problem of Tocharian origins has been resolved over the past century. We know that the Tocharians ended up in the Tarim Basin, so we have one of two choices. We can admit our ignorance, draw a ring around Tocharian territory and pretend that they grew out of the ground while all the other Indo-European groups arrived in their historical homes on the tips of arrows. This, of course, would explain nothing about the origins of the Tocharians. Alternatively, we can draw an arrow with our red felt-tipped pen. We know where to put the point—in the Tarim Basin. But where do we place the base of the shaft of the arrow? Where do we set it into flight? So far this depends solely; it would seem, on where one finds the closest linguistic neighbours. If this is, for example, Celtic, then we have an incredibly long arrow that stretches from Central Europe all the way to the Tarim Basin (generally this is held to be visually too ‘extreme’ and so the flight of the arrow is frequently truncated: we only get a shorter arrow arching its way out of Central Europe in an easterly direction). There are no cultural or archaeological arguments for these arrows; they represent felt-tipped pen archaeology.

A Brief History of Tocharian Time

Now those who have sought to employ dialectal arguments to position the pre-Tocharians in space have often (though not always) ignored the temporal problems involved in any such claims. For example, if one were to accept that there was a particularly close association between Tocharian and Celtic or Germanic, what does that mean in terms of both real time and real space? Invariably, Tocharian seems to finish a poor second and the linguist begins to draw his or her arrow of Tocharian migrations from the eastern margin of whatever is lucky enough to be chosen as its partner. It is with very few exceptions the Tocharians who are required to take the long walk (or ride) while the Celts (or Germans) get to sit in their historically attested seats of central and western Europe. Why? As the relationships under discussion supposedly occurred before the formation of Proto-Celtic and Proto-Tocharian, neither need have achieved their historical seats at the time of their proposed relationship.

Knowing when such relationships should have developed is absolutely critical since space, at least in our four-dimensional universe, has no meaning without time. And here we must face the frequently ignored asymmetry of the Indo-European space-time continuum. The dates that have been estimated for the emergence of the various European proto-languages, i.e., Celtic, Italic, Germanic, Baltic and Slavic, all of which have been partnered with Tocharian, is generally set broadly within the period c. 1500–500 BC (we have also seen that Tocharian might be dated to this period). Before this time we seem to be dealing with some ill-defined period of Late Indo-European during which pre-Celtic, pre-Germanic, etc., linguistic tendencies were coming into being. This would not be of such great importance if it were not also apparent that Greek, Anatolian and the Indo-
Tocharian and northwest European, say in the 1st millennium BC, then this would certainly drag the pre-Tocharians much farther west since we can be fairly confident of where the immediate predecessors of the northwest European proto-languages were sitting after 1000 BC. If we accept some form of northwest orientation for Tocharian, can we provide any evidence for its date?

Strictly speaking, the answer is probably 'no' but there is some circumstantial evidence to suggest that any putative connections must be earlier rather than later. There is considerable evidence that Germanic, Baltic and Slavic once formed some type of linguistic continuum that experienced innovations in both morphology and vocabulary. The latter offers some vague indices of dating since it includes cultural terms such as 'gold', 'silver', 'rye' and 'oats', all of which are unlikely to have been adopted across northwestern Europe prior to c. 2000 BC.

Tocharian participates in none of this particular shared cultural vocabulary (we have no idea what, if any, were the Tocharian words for 'rye' and 'oats' but we do know the words for 'gold' and 'silver') and so if it indeed does share certain unique forms with this area, it is more likely to have been earlier when the geographical implications of such a relationship would be ambiguous at best.

Those who have attempted to associate Tocharian with other Indo-European language stocks appear to be in agreement that there are no grounds whatsoever for seeing a special genetic relationship with Indo-Iranian. Taken in the light of the probable date of loanwords from the Indo-Iranian languages, this provides us with some interesting parameters for locating the ancestors of the Tocharians in both space and time. We generally imagine that the Indo-Iranian language stock formed sometime around 2500-2000 BC somewhere in the east European or west Asiatic steppes, i.e. between the Ukraine and western Kazakhstan; and wherever the ancestors of the Tocharian language lived, they should not have been part of the nucleus of the Indo-Iranian at this time.

Those who argue that the Tocharians cannot be closely associated with any other Indo-European stock would appear to be noncommittal about the geographical position of the pre-Tocharians except for the spatial constraints required by no evident relationships, i.e. where we find the other Indo-European groups would not appear to be the best place to look for the pre-Tocharians. Implicit in this is the belief that pre-Tocharian was separated from the Indo-European linguistic continuum at a quite early date before other dialectal relationships developed. This, incidentally, does not demand that the Tocharians were necessarily only a small group of speakers who wandered off from the rest; they may have been a major group which, by the time they entered the historical record, had been whittled down to those surviving in the Tarim Basin (in the same way that the distribution of Celtic languages has declined to a few pockets in Atlantic Europe). On the other hand, had they occupied a larger area of Eurasia, one might have expected, after their absorption by other Indo-European stocks, particularly...
Tocharians and Afanasevo

With so many possibilities, it is difficult to make logical demands on where we should expect to find the prehistoric Tocharians. Nevertheless, there are aspects of Tocharian origins that more easily accommodate one sort of solution than others. The type of solution that would best fit the linguistic evidence is one that derives the Proto-Tocharians from somewhere that was not in obvious or close contact with the prehistoric ancestors of the Indo-Iranians and that permitted them to enter their historical seats before the arrival of the Iranians. Such an origin would permit them to evolve linguistically without sharing the innovations, e.g. satenization, experienced by the Balto-Slavic-Indo-Iranians. As this would appear to have reflected a continuum that spanned the region from eastern Europe to Kazakhstan, we might look for the Tocharians on either side of this continuum. In this way, they could develop in isolation before their arrival in the Tarim Basin. There is one archaeological culture that might just fit the bill.

The Andronovo culture, which we have already seen to be widely associated with some stage in the development of the Indo-Iranian languages, was not the first culture to practice stock-raising on the Asian steppe. Before the expansions of Andronovo tribes, there were the Afanasevo communities of the eastern steppe in the region between the Minusinsk Basin of the Yenisei River and the Altai Mountains. Dating to the period c. 3500-2500 BC, the Afanasevo culture represents an archaeological enigma as its closest cultural relations appear to lie 1,500 km (930 miles) to its west in the European steppe.

Most of our evidence derives from burials which were deposited in stone-walled enclosures, rectangular or circular, which might then be capped with stone slabs. The deceased were extended on their backs but with their legs flexed, a rather rare burial position attested primarily on the European steppe in the Yamna (pit-grave) culture. Like the Yamna culture, ochre was also employed in the burials. The pottery included round- and pointed-based vessels, similar in a genetic sense to the pottery of the Yamna culture. Specifically more closely related to the Yamna and later Catacomb cultures of the European steppe were the ceramic 'censers', small legged bowls which are believed to have held some form of aromatic substance (one from Romania contained traces of hemp which we know was later utilized in rituals of the Iron Age steppe nomads). There are also traces of horse (whether wild or domesticated is not determined), metallurgy employing copper and silver, and possibly wheeled vehicles which are depicted on stone slabs. Finally, the physical type of the Afanasevo culture is Caucasian and has been most closely related to that of the European steppe populations.

It is still very difficult to explain the origins of the Afanasevo culture since it appears so far east of its European cousins and there is a stark absence of intermediate sites excepting a handful of Afanasevo-type burials. The current working hypothesis would explain them as one of the earliest food-producing cultures to emerge out of the European steppe which, for reasons that still elude us, pushed far to the east. The territory between the Volga-Ural and the Yenisei is archaeologically poorly known and it is hoped that the filling of the gaps that might underwrite this model will be accomplished eventually.

From a linguistic point of view, the Afanasevo culture provides us with the opportunity to fit a culture to the type of linguistic development that we have suggested for the Tocharians. Here is a culture that was linked to developments in Europe but isolated far to the east. Moreover, when the Andronovo culture emerged, the culture most closely associated with the Indo-Iranians of the steppelands, the Afanasevo culture had already disappeared. If Tocharian documents had been found in the Minusinsk Basin or in the Altai Mountains, we would not have had any great difficulty explaining their origins. But this is not where we discover our written evidence for the Tocharians but rather in the Tarim and Turpan basins.
Although the centre of the Afanasevo culture lies far to the north of the Tarim, there is some evidence that it spilled southwards and Afanasevo sites have been claimed for both Mongolia and northern China. Some suggest that we may look even farther south, right into the heartland of the mummies of East Central Asia.
CHAPTER TEN

Who Were the Mummies?

It should be abundantly evident by now that our attempt to reconstruct the ethnic prehistory of East Central Asia will require quite a feat of both archaeological and linguistic legerdemain. From a period comprising ethnically-anonymous and poorly-known prehistoric cultures, which provide the cultural context for the Xinjiang mummies, we must navigate through a period of ambivalent parahistorical references to populations of East Central Asia dating primarily from the Han and later dynasties and, somehow, tie these into the evidence of the myriad languages of our still more recent written documents. We would do well, then, to keep our assumptions in mind, not so much to guarantee the validity of any of our conclusions as to ensure the integrity of our speculations. Here are a few of the more obvious.

The mummies and the populations from which they were drawn need not have belonged to a single ethnolinguistic group but may have spoken a number of different languages. This possibility is enhanced by the differences between the prehistoric populations with respect to their date of deposition, geographical distribution and human physical type. It is quite possible, then, that there is no single answer to the question: who were the Xinjiang mummies?

Although it is more likely that people from a single cemetery at a single time were themselves linguistically related, this too is an assumption. The nature of the tribal confederations of the steppes suggest considerable linguistic mixing and the documentary evidence of the Tarim oases in the historical period suggests that bilingualism was probably the norm, rather than the exception with multiple language groups occupying any given centre. In short, we are not asking what language the mummies may have spoken but what language they spoke at home?

The language(s) spoken by some or all of the prehistoric mummies may be totally unknown, i.e. did not survive to be reflected in the documentary record of the Tarim and Turpan basins. Instances of the near disappearance of a language are frequent enough to make this a genuine rather than a merely theoretical possibility. From Europe, we have only very meagre written evidence for the Pictish language of Scotland, the Taressian and Libnan languages of Spain and Portugal, Iron Age Balkan languages such as Thracian, Dacian and Illyrian, Etruscan and other non-Latin languages in Italy, languages from along the western coasts of Anatolia, etc. In all these cases and many others, had the written record arrived more than a few centuries later than it did, we might well have been deprived entirely of the evidence that such languages were once spoken in their respective regions. In some cases, traces of an extinct language are also recoverable from the place names of a region, e.g. Gaulish names underlie the names of many a French town. But here in East Central Asia, the recovery of earlier linguistic horizons is likely to be extremely hazardous since it is often difficult enough recognizing 'Western' personal or place names in their Chinese transcriptions, much less reconstructing the phonetic shapes of non-Chinese place names purely from their citation in early Chinese texts. The great scholar of Saka studies, Harold Bailey, listed a large number of personal names from the Khotanese documents which did not appear to him to be Iranian, Chinese or Tibetan in origin, Douglas Adams examined them and added Tocharian to the list. If Khotanese phonological changes have not totally obscured loanwords from more familiar languages, it is possible that we may be dealing with some unknown substrate language (or the elusive Burushaski?) in the southwest part of the Tarim Basin.

Finally we have made it clear that the mummies themselves are only the accidental residue of the prehistoric inhabitants of East Central Asia. Pinning an ethnolinguistic name on any particular group of mummies is by no means the same as reconstructing the complete prehistory of the populations of ancient Xinjiang. For example, our evidence for mummies along the northern Tarim Basin is negligible, yet this would become the major northern route of the Silk Road. On the other hand, attempting to assign specific ethnic identities to the mummies is probably as good an introduction as any to the wider problem of reconstructing the ethnic prehistory of East Central Asia.

Identifying the Mummies

Focusing our attention on those prehistoric mummies whose physical type has been labelled 'Caucasoid', we can divide them into four broad groups:

1. The Chačšan group comprises those mummies so far recovered from the cemetery at Zaghunluq.
2. The Lopnur group provides the earliest and most diverse assemblage of mummies from the excavations of Aurel Stein, Sven Hedin, Folke Bergman and the more recent excavations at Qawrigulu and the Tawin River. All these fall within the general region of the ancient kingdom of Kroran.
3. The Qumul group comprises the mummies found near Qumul (Hami) east of the Turfan Basin.
4. The Turfan group is represented by the cemetery of Subeshi (here also are the later Mongoloid mummies from Astana).

We shall not discuss here the recently unearthed mummies from the south-central and southwestern sites of Niya and cemeteries such as Sampul in the vicinity of Khotan, even though they are Caucasian, because many of them date after the Early Iron Age (post Han dynasty), our cut-off point for trying to understand the movements and positioning of the peoples of East Central Asia; most of them have not been adequately conserved or even preserved, but have been left in pieces in their original sites (this is sadly often true of mummies discovered elsewhere in the Tarim Basin and surrounding areas); and of these few have been brought from the field into the museums or archaeological research units or have yet been made available for study.

Can we blithely leap from the distribution of the mummies to the later evidence of language? Let us consider what would happen if we attempted to assign an ethno-linguistic identity to a Bronze Age burial in England. Extrapolation would lead us to identify it as English (or Anglo-Saxon) which, as we know with historical hindsight, it simply could not be. Are we likely to make the same type of mistake if we adopt this approach in assigning identities to the Tarim mummies? It will be up to the reader to judge the validity of our logic.

From a purely geographical and temporal perspective, the Lopnor (Krorian) group falls within a territory that later provides us with Kharoshthi Prakrit as its administrative language. We can be certain that this was not the language of its pre-Han inhabitants because this particular Prakrit language, derived from northwest India, did not even exist until the latter part of the 1st millennium BC, and reflects a relatively late phenomenon of the region, associated with the spread of Buddhism and Buddhist monks no earlier than the first centuries BC and presumably somewhat later. And we have also seen that the Prakrit documents retain traces of another, presumably earlier substrate language that might be termed Kroranian (it includes Saka loanwords as well). On the basis of its personal names, some of its vocabulary, and principally its grammatical forms, we have seen in the last chapter how this Kroranian may be set alongside the other Tocharian languages as 'Tocharian C'. Since we find ourselves faced with but two choices – Prakrit or Kroranian – as the language of the earliest mummies of this region, then Kroranian is the clear favourite.

While Kroranian was apparently the earlier vernacular language of Kroran, the fact that the majority of the documents with Tocharian C material derived from the site of Niya far to the west may indicate that much of the southeastern Silk Road, at least that between Niya and Kroran, was once occupied by Tocharian C-speaking populations as well. If one accepts this then the simple observation that the Charchan group is situated between Niya and Kroran invites us to assign these mummies also to Tocharian C.

But before we can assign our first two groups of mummies to the Proto-Tocharians, we should recall the evidence of Chinese physical anthropologists who propose the movement of at least two Caucasian physical types into the region: an earlier Proto-Europoid at Qiayqghul and a later Indo-Afghan type at Kroran. Both of these are earlier than our earliest Prakrit documents from the region which date to about the 3rd century AD. If we accept these divisions can we say which, if any, of these two 'waves' spoke Kroranian, the language of the Tocharian C substratum? Any attempt to secure a tight relationship between physical type and language is highly suspect and here we will regard physical type as merely circumstantial evidence for geography rather than direct evidence for a particular language. We know that the direction from which the Prakrit language must have come was the west, i.e. northern India via Bactria. As the Saka were widespread across the steppe and also penetrated Bactria and even northern India, they
could enter the Tarim basin from either the west or the north. Of these two potential sources the west is surely the preferred option because the main source of Saka texts is Khurstan on the southern Silk Road, which was not readily approachable directly from the north other than along a desert road. We have also seen that the physical evidence along the entire southern route of the Silk Road back into Bactria appears to be marked by the same general physical type. The trajectory of this Indo-Afghan type, therefore, is most easily assigned to those who carried Indic or Iranian (Saka) languages from the west, and this is further confirmed by the survival of Saka’s closest linguistic relations in the Pamirs. Thus, by a crude process of elimination, this leaves the earlier attested Proto-Europoids (or Hemphill’s peripheral Qawrighul people) as the more likely candidates to be identified with the Tocharian C substrate. There is, admittedly, a chronological gap between the latest mummies and Khotanian texts but we should recall that the mummies from, for example, Zaghunluq (‘Ur-David’), the baby in the blue bonnet, the tripartite woman, the ‘Scream Baby’, etc. date to c. 600 BC and we are only suggesting that they be associated with a linguistic substrate that must have been established before the 3rd century AD. As Tocharian loanwords are attested in Chinese a few centuries before the Khotanian documents, we can perhaps narrow the gap still further between the prehistoric mummies and the Tocharian languages.

The Turpan group with its burials dating to about the 5th or 4th centuries BC falls within a territory where historical evidence reveals both Tocharian A and B documents. While it is difficult to know how far back we can extrapolate either of these languages as the vernacular language of the region, Tocharian at least supplies a plausible candidate for the Turpan group of mummies since we have no other language in this region that offers a credible alternative. In terms of general physical type, here too the Proto-Europoid type appears but alongside the Indo-Afghan type, with the Pamir-Fergana type apparently the most recent.

Finally, although the Qumul mummies lie beyond our documentary evidence for the vernacular languages of East Central Asia, we still have a few hints as to what language they may have spoken. We do, for example, have Tocharian documents a mere 150 km (93 miles) to the west. We might also now recall Werner Winter’s suggestion that the language which emerged in written sources as Tocharian A may have been spoken farther to the northeast, i.e. in the direction of the later Turkic penetration of East Central Asia, before it had been taken up as a ‘dead’ literary language. When we examine the general physical type of the more easterly populations such as those from Yarimqala, it is the same as that found at Qiwrighul. The chronological gap between the Qumul mummies, dated to about 1000 BC, and Tocharian texts of about the 7th century AD is indeed great but here we may recall that with respect to Tocharian A we are dealing with an already dead language which should have preceded its historical attestation in this region by a matter of centuries, thus narrowing the gap between the mummies and our language horizon. Here too then one might propose a north–south movement of early Tocharians.

The evidence of a crude slash from Occam’s razor would seem to suggest that the majority of our earliest prehistoric mummies may be regarded as ancestors of the later Tocharians. These conclusions should, however, be tempered with the observation that, except on the basis of medieval wall paintings, we cannot speak of the physical type of those populations who occupied the territory most closely associated with our Tocharian B documents, i.e. around Kucha. In other words, we cannot easily relate the Bronze Age populations with the later oasis states mentioned in Han documents that produced the Tocharian documents of the historical period. What we have proposed is merely that on the evidence of physical typology and geographical distribution, the Tocharians would appear to derive from the north and the Indo-Iranians largely from the west: how robust are such conclusions when we add into the equation the evidence of archaeology?

### The First Farmers

We have seen how our earliest archeological evidence for human occupation of East Central Asia is extremely meagre and it is impossible to determine with any certainty from which direction the Tarim and neighbouring regions were initially settled if colonization preceded the Bronze Age, i.e. c. 2000 BC. From the beginnings of the Bronze Age itself the circumstantial evidence that we have surveyed would render it more likely that it was settled from either the north or the west rather than the south or east. The paucity of a Mongolid physical type among the earliest human remains and its confinement to the far east of East Central Asia suggests that Mongolid populations entered late and from the east and southeast. Furthermore, we have a good idea of the settlements and cemeteries of the Neolithic Yangshao culture of northern China which expanded westwards into neighbouring

<table>
<thead>
<tr>
<th>Mummy groups, languages and background physical types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>Charchan</td>
</tr>
<tr>
<td>Lopnor</td>
</tr>
<tr>
<td>Qumul</td>
</tr>
<tr>
<td>Turpan</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Gansu and yet we lack any such evidence for typically Chinese Neolithic settlements and cemeteries in East Central Asia. Nor, given our knowledge of the subsistence economies of the early farmers of northern China, is it likely that they would have found settlement east of the Jade Gate viable, much less attractive.

If the east is excluded, what about the southwest, i.e. greater India? Here we have a Neolithic culture in Kashmir by the middle of the 3rd millennium BC which itself appears unrelated to developments elsewhere in India. But other than the occasional employment of ochre in burials and the later adoption of mud brick in building, there is very little archaeological evidence to tie Kashmir with the earliest inhabitants of the Tarim Basin or those who introduced the agricultural economy to the region.

The earliest Bronze Age cultural remains are still assigned to the Qiwrghul culture which was situated near Lopnor. All the other earlier Bronze Age cultures (i.e. Yanbulaq, Ayding Lake, Yongdala (Xintala)) are localized in the northeast or northwest (e.g. Qaradong (Haladan)) of the Tarim Basin. These are the cultures that provide the initial context for the Xinjiang mummies and if we can exclude the east and south as their probable area of origin, we must then face the two most probable archaeological solutions to the origins of the Xinjiang mummies – the western (Bactrian) and northern (Steppes) hypotheses.

The Bactrian Solution

Early farming villages appeared southeast of the Caspian by the 6th millennium BC and began a gradual expansion eastwards. By c. 2000 BC we have seen how the Bactria-Margiana Archaeological Complex (BMAC) was employing irrigation agriculture to exploit the oases of Margiana and Bactria, the latter of which had not seen the introduction of agriculture (other than that associated with the site of Shortugai in the northeast of Bactria which had served as a trade outpost of the Harappan civilization) prior to the BMAC. One might then be tempted to propose that the BMAC not only spread subsistence agriculture into Bactria but that its farmers continued much farther east, perhaps through Ferghana, to enter the Tarim Basin. There are a number of factors in favour of this proposition.

Since the earliest agriculturalists of the Qiwrghul culture do not seem to appear until after 2000 BC, a time after which the BMAC had already formed and begun to expand, there is no great temporal difficulty in deriving agriculture from the BMAC, although the distance involved is formidable and so far unbridged by intermediate sites. Also, western crops such as wheat, and livestock such as sheep entered China from the west and both of these were possessed by the BMAC. More importantly, the BMAC had successfully mastered the techniques of exploiting oases with irrigation agriculture and so it would have possessed the technological prerequisites for exploiting the particular environment of the Tarim Basin.

In addition to its subsistence basis, there are a few other features in favour of the BMAC. We have mentioned how its use of mud brick in domestic architecture finds some parallels in the Yanbulaq culture where we also found mud brick in both the tombs and domestic architecture. And we have also seen how the BMAC reveals striking evidence for the consumption of stimulants and hallucinogens in religious rituals, i.e. the ritual preparation of poppy, hemp and ephedra. It is tempting to tie this to burials in the Lopnor region that reveal the widespread use of ephedra. Finally, the BMAC provides a strong candidate for the culture of early Indo-Iranians and, therefore, can be related to one of the major language groups that we encounter in the Tarim Basin.

A case, then, can be made for an extension of West Central Asian culture eastwards into the Tarim Basin. But the evidence is still hard to compelling. While BMAC burials may be found far from Bactria-Margiana in, for example, Baluchistan, there are no actual BMAC burials nor ceramics in East Central Asia. On the other hand, the successful exploitation of oases and the use of mud brick (absent from neighbouring Bronze Age Gansu) do suggest western contacts of some form. But these are seen most clearly in the Yanbulaq culture rather than the Qiwrghul culture which still appears to be the earliest of the East Central Asian Bronze Age cultures. Moreover, despite the presence of West Central Asian 'influences', Chinese archaeologists attribute much of the Yanbulaq culture to local populations who employed a material culture similar to that of the Qiwrghul culture. Finally, the evidence of the human physical type at Yanbulaq, i.e. primarily Mongoloid but with some Proto-Europoids, is incongruent with the type of
populations we find in Bronze Age Bactria-Margiana where a more gracile form of Caucasian predominates. It is difficult then to imagine that the human vector that provides us with our earliest Bronze Age populations, including most of our prehistoric mummified remains, came from Bactria.

The Steppe Solution

The second major hypothesis emphasizes connections between Qawrighul and other early cultures with those of the steppelands to the north (and northwest), specifically the Afanasevo and Andronovo cultures. It might be noted that contacts with the steppelands were not a single event but a protracted process which did not cease until the Uyghur conquest of East Central Asia. What else does the Steppe hypothesis have going for it?

As with the Bactrian hypothesis, the steppelands offer evidence for the spread of a mixed farming economy into the Minusinsk Basin and the Altai Mountains in the 3rd millennium BC (Afanasevo culture), dates that precede the earliest appearance of the farming economy in the Tarim Basin. We have also seen how the important centre of bronze metallurgy closest to East Central Asia was situated in the steppe zone and there was an unquestionable penetration into East Central Asia of bronzes of the Andronovo type throughout the 2nd millennium BC.

While we have no evidence of the use of ephedra among the steppe tribes, we have already seen that they did share in the cultic use of hemp, a practice that began from Romania east to the Yenisei River from about the 3rd millennium BC onwards where it was later encountered in the form of apparatus for smoking hemp found at Pazyryk.

Elena Kuzmina, one of the foremost authorities on the peoples of the eastern steppelands, has noted a number of specific similarities between the Qawrighul and steppe (Afanasevo, Andronovo and later) burials:

a) use of enclosures such as fences, stone circles and radial features;

b) use of timber or bark roofs (or floors) for chambers, often lined with felt or pelts on the inside;

c) inclusion of animal sacrifice with burial;

d) use of ochre in burials;

e) similar clothing, i.e. leather boots, felt caps.

As with the Bactrian hypothesis, the evidence does not permit a firm conclusion in the sense that we can describe the Qawrighul culture as a regional variant of the Afanasevo or Andronovo culture. Ceramics, which were abundant in both

the Bactrian and steppe burials, were not found in those of the Qawrighul culture (we have no idea what type of pottery they employed or whether they even utilized ceramics – the lack of local clays may have largely precluded their manufacture in some of the southern sites) nor did the Qawrighul culture reproduce the stone enclosures that surrounded the Afanasevo burials.

If the Afnasuevo or early Andronovo culture had penetrated southwards into Xinjiang, we should expect to find their monuments initially in the Junggar (Yarish or Dzungarian) Basin which was midway both in geography and environment, between the Yenisei-Alta and the Tarim Basin. Do we have an intermediate culture in this region? Here we recall the remains of the Keremchi (Ke’ermuq) culture which was situated along the northern foothills of the Junggar Basin. On the debit side, this culture lacks radiocarbon dates and most of its tombs have been plundered so its precise chronological position remains highly problematic. On the plus side, however, it may offer some evidence for a 'missing link' between the Afanasevo and Qawrighul cultures. The Keremchi cemeteries employ stone-built enclosures, erect anthropomorphic stone figures, and exhibit round-based primitive clay vessels decorated with incised ornament that are reminiscent of Afanasevo pottery. There are also ceramic parallels with the early Andronovo culture. Without precise dates for this culture, conclusions must be speculative and, as was the case with the Qawrighul culture, we may be dealing with a very long-lived phenomenon in the first place.

162. Hemp-smoking apparatus from Pazyryk; the tripod held up a tent in which one inhaled smoke from being heated on a brazier.
Nevertheless, if the similarities with the Afanasevo and subsequent cultures can be translated into dates, the evidence provides additional support for the expansion of a steppe-derived culture into East Central Asia by the 2nd millennium BC. Kermechi burials are found also in the southern Altai and as far south as Ürümqi and the Yanbulaq culture.

**Splitting the Differences**

We must reconcile ourselves to the fact that the earliest Bronze Age culture, possibly the earliest farmers, in East Central Asia were representatives neither of the Afanasevo culture nor of the BMAC per se. Of the two cultures, the Afanasevo boasts earlier dates and more mobile antecedents. So far it reflects the earliest evidence for the exploitation of domesticated animals in the steppelands far east of the Ural. To connect the earliest Bronze Age settlement of East Central Asia with the Afanasevo culture would seem to require one of the following models.

One might propose that there was an earlier common source for both the Afanasevo and Qawrighul cultures that spread across Kazakhstan eastwards to the Yenisei and southeastwards into the Tarim Basin. How early such a phenomenon might be is constrained by the fact that the earliest Afanasevo dates would carry this culture back to c. 3500 BC (already embarrassingly early given its supposed antecedents in the west) and so we would have to envisage a spread of steppe-keepers eastwards during the first half of the 4th millennium BC. This is not impossible but evidence for such an agricultural expansion is not supported by archaeological data and we know of cultures east of the Ural that date to the 4th millennium BC that are clearly non-agricultural. Unless the archaeological picture of Kazakhstan alters radically (a possibility since there is still so much unknown) an earlier common source for both the Afanasevo and Qawrighul cultures is probably not the preferred option.

Alternatively, the spread of agriculture and stockbreeding to the east which is indicated by the Afanasevo culture may have also seen a subsequent movement south into East Central Asia. If this were the case, then populations such as the Qawrighul culture would either have had to have abandoned the deposition of pottery in graves (as is well known in the Afanasevo and later Andronovo cultures) or possibly to have neglected the manufacture of pottery altogether and replaced it by organic containers (basketry) highly developed at Qawrighul. This latter situation is not unheard of and is encountered, for example, at the transition to the Iron Age in Ireland and western Britain. But neither settlement of the steppelands of the Minusinsk Basin nor the uplands of the Altai should have prepared such newcomers for the environmental regime of the Tarim Basin; how could mixed steppe-keepers and farmers from the north have successfully developed the irrigation agriculture required to exploit the oases of East Central Asia?

It is this need to develop an irrigation-based economy that turns our attention westwards back to West Central Asia, specifically Bactria and Margiana, where we have seen that a very successful oasis economy was developed by c. 3000 BC. But we have also seen that so far there is no question of a direct import of the BMAC into the Tarim Basin and the parallels tend to be general rather than specific. The case for external influences improves markedly between the Tarim Basin and the west, specifically between the Tarim Basin and Ferghana, in the later 2nd millennium BC, a time after the initial settlement of the Tarim by Bronze Age farmers. For this reason we have an incongruity between a burial rite and material culture that may be derived from the steppelands and an irrigation economy that should derive from the agricultural oases of Central Asia. How can these two very different components be reconciled?

Our most obvious recourse is the structure of relationships between the steppes populations and those of the oasis communities. We have already seen in West Central Asia that steppe tribes could settle and adopt irrigation agriculture. But in West Central Asia, where we witness such phenomena, this symbiosis always occurs where we have prior agricultural settlement. This is not a model that we can easily transfer to East Central Asia since we have no substantial evidence for agricultural settlements prior to the establishment of the Bronze Age, nor can we demonstrate that the physical type associated with such an intrusion of farmers predates those from the northern steppelands. To prepare a 'northern steppe culture' for life in the Tarim Basin, we might expect that it first came into contact with one of the existing West Central Asian oasis-based cultures before it arrived in East Central Asia. Some linguistic evidence may support just such a hypothesis.
The Linguistic Stratigraphy of Tocharian

We have seen in the preceding sections that one of the critical issues is the correct ordering of our evidence, be it that of populations or archaeological influences. The problem before us now is arranging in correct chronological order the different prehistoric languages spoken in the Tarim Basin. Various degrees of difficulty accompany such an exercise but from our perspective the primary goal is to arrange correctly the relationships between the Tocharian and Iranian languages as these are the languages most likely to have served as the vernaculars of its prehistoric populations.

As we have already seen, every language contains something of its own cultural history in its vocabulary. Some of the vocabulary will be inherited from Proto-Indo-European, some will be new words created from older inherited elements. These do not hold the same interest for us here that loanwords do, since borrowed vocabulary, particularly diachronic cultural vocabulary, may point to the time and place of foreign contacts.

As we have remarked before, Chinese appears to have had a very minimal impact on the Tocharian language. To be sure, we have the odd loanword, e.g. Tocharian klu ‘rice’ from Old Sinitic *gliow (modern Chinese dào) and names for units of measurement, e.g. sunk, tǎow (≈ 10 sunk) and cǎk (≈ 10 tǎow), all transparently from Chinese. But the impact here is minimal, culturally predictable and comparatively recent. Otherwise, there is no evidence that the Tocharians gained their vocabulary for the native flora and fauna of the Tarim Basin from earlier Chinese inhabitants, nor do they appear to have gained their agricultural vocabulary from this source. This lends additional support to the argument that the Chinese were both latecomers into the Tarim Basin and did not have a significant impact until relatively late in the easternmost regions.

Now many of the other loanwords found in Tocharian are connected with the religious and social life associated with Buddhism and hence the sources here are Buddhist Hybrid Sanskrit, Prakrit and other Indic languages, occasionally perhaps filtered through Iranian intermediaries. These may be easily enough associated with the relatively late spread of Buddhism among already established populations, both Tocharian and Iranian-speaking.

By far the greatest number of loanwords in Tocharian would appear to derive from Saka or some other East Iranian language and these fill out a larger semantic range. In some instances they appear to have been loanwords associated with exchange, e.g. Tocharian B pita ‘price’ may derive from Saka pīhā ‘price’, and Tocharian A ārət, Tocharian B peri ‘debt’ would be derived from Saka pīra- ‘that which is to be paid’. There are other words which might also be assigned to the general realm of commerce, e.g. units of weight and measure. Iranian military terms also had an impact on Tocharian and when one considers depictions in Buddhist temples of Tocharian knights kitted out like Sassanians with long swords, one might recall that the Tocharian B word for sword, kertte, is derived from some East Iranian language, e.g. Avestan karota- ‘dagger’. Bactrian also supplied loanwords filling out similar political contexts, e.g. kaminda ‘head’, ‘chief’ (god) gives Tocharian A kāktamārkt ‘ruler’. It is also important to note that the Tocharian word for ‘iron’ (Tocharian eriça or ṣecca) would appear to be related in some way to Iranian (Ossetic). In some instances the loanwords are connected with exotic animals, e.g. Tocharian B mewiy (‘tiger’ from Saka maṣav-, ‘tiger’ or Tocharian B ekiṣnek ‘pigeon’ which seems related to Saka asačak ‘pigeon’. By and large, however, it would be exceedingly difficult linguistically to make any case for the temporal priority of Saka in the Tarim Basin with respect to Tocharian. Of the approximately 25 or 30 loanwords, most could be easily explained as later borrowings passed between communities along the Silk Road connected with the rise of urban or Buddhist institutions. Most importantly, there is no evidence that the vocabulary of agriculture in the Tarim Basin specifically derives from Saka.

Asses, Canals and Bricks

Although Saka and Sogdian are the earliest of the Iranian languages directly attested in the Tarim Basin, there are some Iranian loanwords in Tocharian that may derive from a still earlier period. Two of these have only recently come to light in the form of a Tocharian B text that documents the transfer of an estate. Within the text there is reference to the orotṣa mewiy ‘great canal’ and the ārē, some form of watercourse that can serve as an estate boundary. Douglas Adams has recently suggested that both of these words are early loans from an East Iranian language (or proto-language) and, while there are a number of ways to explain how these loans could have taken place, only one is not problematic. As irrigation agriculture is the only type of agriculture that one can practise in the Tarim Basin, it is difficult to see how the Tocharians could have eked out a living there before they knew of irrigation. It is also unlikely that the Iranians and Tocharians both arrived at precisely the same time, the Iranians passing onto the Tocharians the technology of agriculture before the latter had starved to death. The simplest explanation would be to place the Iranians in position first and then have the Tocharians, largely pastoralists, wander in and adopt agriculture; this is precisely the type of pattern we have earlier encountered in West Central Asia where steppe pastoralists moved into the oases and adopted the agricultural techniques of the earlier inhabitants. But, as we have already seen, we do not really have typically West Central Asian BMAC farmers occupying the Tarim Basin, nor does our review of the physical evidence suggest that populations from West Central Asia arrived in the Tarim and Turpan basins earlier than those from the north. None of our usual models will do.

Adams suggests that the most convenient explanation would involve the adoption of irrigation techniques and East Iranian terminology by the
Tocharians en route to their historical seats. This model would fit nicely for those archaeologists who have noted that while there is no evidence for the BMAC in East Central Asia, certain similarities between the Yambulak culture and that of the BMAC do suggest some form of mediating culture, presumably a more mobile pastoral culture of the Andronovo type. We may suggest then that Tocharian populations moved from the steppe, through the Altai and Tängri Tagh, and south into the Junggar, Tarim and Turpan basins, settling in the oases of the latter two to engage in irrigation agriculture.

It might be objected that no one entering the Tarim and Turpan basins who had to make their way through the surrounding mountain passes is likely to have carried the techniques and vocabulary of irrigation agriculture with them. This is not a serious objection as the underlying East Iranian word for 'irrigation canal' that we find borrowed into Tocharian may also be found in mountainous regions where it is known in Sorikoli, one of the Pamir languages related to Saka, and cognates are also known in the far northwest of India in the Nuristani and Dardic languages. That common words connected with irrigation can be found on the western mountainous approaches to East Central Asia at least suggests the possibility that Tocharian speakers passing through this region could have adopted such terms and techniques. Of course, a western entry would not really solve the problem before us if East Iranians were in occupation in Fergana and perhaps present in the Altai as well, it might provide a plausible route for northern steppe peoples southwards into East Central Asia.

Another instructive piece of evidence is the Tocharian word for the ass, Tocharian B kesaɨpo which is universally agreed to be related to Old Indic gorâbha. If this is a borrowing, it is a very early one when the underlying form in Proto-Indic and Proto-Tocharian was something like ‘gorâbha’. Now such a form could just as well be Proto-Indo-European and so we are not dealing with some word borrowed from the period of Buddhist expansion in East Central Asia; a word ‘gorâbha’ should have been circulating among peoples in the 2nd millennium BC or earlier. We are uncertain as to whether the original word referred to the domestic ass or the wild onager, and was then transferred to the domestic ass. The ass was originally domesticated in North Africa and then expanded into Mesopotamia by the 4th millennium BC, spreading from there both northwards and eastwards; its first appearance in northwest India is around 2000 BC. On the other hand, the so-called ‘half ass’ or onager ranged across the entire steppes and south into Central Asia and northern India. Whatever the original meaning of the word, the linguistic evidence suggests some form of contact between Tocharians and Indo-Iranians long before the opening of the Silk Road and the spread of Buddhism.

We have already seen that one of the cultural features that tends to link the prehistory of East Central Asia with that of West Central Asia is to be seen in the use of mud bricks in domestic and funerary architecture. Of interest then is the fact that Tocharian ēkem ‘clay’ would appear to be related to Avestan ēspā ‘brick’, zâmēsēna ‘clay-tile’, Old Indic ēstak ‘brick’, and Khuras (a northeast Indo-Aryan language) ēbā ‘sun-dried brick’, ‘large clod of earth’. This would again point to either a loan from a very early stage of Indo-Iranian or the mutual inheritance by both language stocks of a common eastern Indo-European proto-form ‘isti- ‘clay’, ‘brick’.

So far the linguistic evidence is hardly overwhelming, but it does hint at contacts between Tocharians and Indo-Iranians prior to the spread of Buddhism in East Central Asia. These contacts need not have been particularly intense and we would be giving a very mistaken impression if we were to portray the Proto-Tocharians as wandering nomads who only learned their agriculture from settled Proto-Indo-Iranians. To be sure, they entered the Tarim Basin with their livestock whose Indo-European names were retained, e.g. Tocharian B ƙeːn, ƙas, ƙaːn and sutsə are cognate with the English words ‘cow’, ‘ox’, ‘ewe’ and ‘sow’ respectively. They also retained at least one of the inherited Indo-European words for ‘grain’ (Tocharian B ṭaːn from a late Proto-Indo-European *dhūb, pēh₂-r-, and words for the basic agricultural pursuits of ploughing, sowing, threshing and grinding. The ancestors of the Tocharians were already mixed farmers (agro-pastoralists) and were not ignorant of cereals before they entered the Tarim Basin, nor did they require contact with Indo-Iranians to learn about farming. One curious word, Tocharian B ṭaːn ‘some form of bread’ is possibly related to Indo-Iranian words for ‘wheat’ (e.g. Avestan ganum, Khhotanese Saka ganama) and Hitite kant- ‘wheat’. Although this may have been a common word in the eastern Indo-European world, it is generally taken to be a loanword from some Near Eastern language that spread through Asia. Finally, one other feature should be noted; there is no evidence of the Indo-Iranian languages adopting Tocharian words; our loanwords only seem to go in one direction.

So far then we have found recurring evidence that north = steppe = (?)Proto-Europoid (if one follows the typological school) = Tocharian; and west = oasis = Indo-Afghan/Pamir-Ferganoid (or Hemphill's Bactrian-Alwighul-Kroän group) = Indo-Iranian. Before we put ourselves on the back it is about time we get a grip on things and have an 'assumption alert'.

The logic of all this hangs on our identification of the northern invaders as Proto-Tocharians and not some form of Indo-Iranians. But we have already seen that the BMAC is not the only culture to which we assign an early Indo-Iranian identity; the Andronovo culture and some of its western neighbours on the European steppes are also widely regarded as Indo-Iranian. Indeed, we have given strong reasons to believe that the Indo-Iranians were originally steppe peoples themselves who came to dominate the oasis culture of West Central Asia. And we have also seen that the Qawhul physical type might be related to that of the Andronovo culture. One is at liberty to propose this equation: north = steppe (Andronovo) = Proto-Europoid = Indo-Iranian. The Occam's razor we employed earlier to
separate Tocharians and Indo-Iranians has suddenly become a very dull instrument indeed. It is time we played the Afanasevo card.

The Afanasevo Card: The Short Trek

We have worked our way into a logical corner but at least we have some notion of what type of device it is going to take to extract us from our predicament. We have the Tocharians in the northern and, accepting the Krotarian evidence, southeastern part of the Tarim Basin by the first half of the 1st millennium AD. We have set our arguments to indicate that they should have been there at least by the 1st millennium BC (before the spread of Buddhism) and we have seen why it is easier to derive them from the north than from any other direction (although we must admit that Iranians might also have come from this same direction). Furthermore, we know from our examination of the separation of the various Indo-European stocks that there are no particular reasons to associate Tocharian genetically more closely with Indo-Iranian than with any other stock. We have already reviewed the linguistic evidence and seen that opinion tends to be split between assigning Tocharian its closest relations among various European stocks (Germanic, Greek, etc.) and arguing that it separated from the other Indo-European stocks at such an early date that there are no grounds to presume it was a close neighbor of any other Indo-European group during a large part of its early development. Either way, the Tocharians are clearly not part of the greater Indo-Iranian superstock and so, for linguistic reasons, we will prefer any model that keeps some distance between them and the other stocks throughout a considerable part of their early evolution. On the other hand, they did come into contact with the Indo-Iranians when they had already emerged as an independent stock, presumably sometime after 2500–2000 BC. It is most likely that it was during the early part of the 2nd millennium BC that Proto-Tocharians came into contact with already settled Indo-Iranians and borrowed a few terms and techniques relating to irrigation agriculture, brickwork and possibly the domestic ass.

From an archaeological perspective, one way to accommodate all these demands is to regard the Proto-Tocharians as an offshoot of the Afanasevo culture of the Altai-Yenisei region. As we have already seen, it displays genetic connections with the cultures of the European steppelands (and hence might represent the eastern extreme of an Indo-European linguistic continuum), but it was also isolated from the other steppe cultures with the later Andronovo culture filling the gap between the European steppe and the Yenisei River and then replacing it. If an offshoot of the Afanasevo culture moved southwards into East Central Asia in the centuries around 2000 BC, it could hardly have avoided contact with Indo-Iranians of some sort just to its west and it may thus have adopted the rudiments and some of the vocabulary of irrigation agriculture from them. It would then have moved into the northern and southeastern Tarim Basins where it established itself, with marked cultural change and adaptations, and continued until the historical expansion of the Uighurs.

In this model the Proto-Tocharians moved south from the Altai region in about 2000 BC to settle in the northern Tarim and the southeast (Krotarian). Subsequent movements of populations from the north may have carried Iranian speakers into the Tarim Basin as well but they never achieved linguistic ascendancy in the north and east. The Saka, however, entered the Tarim from the west during the 1st millennium BC and established themselves in the northwest of the Tarim and in the south at Khotan. How does such a model work geographically?

If we look at the map of the approaches to the Tarim Basin then the natural route would be from the northwest towards the southeast, i.e. from territories that may have once been part of the Indo-Iranian chain of languages of West Central Asia to the Altai southeast into the land of our earliest mummies. To this day, the best passes for entering the Junggar Basin from the north are along its northwest side where one can skirt the forbidding centre of the basin and move along the southwestern mountain slopes to the south. This would bring northern immigrants through a funnel
that discharges near Ürümchi and the major pass south through the Tangri Tagh into the Turpan Basin and then the eastern portion of the Tarim Basin. Potential immigrants from the Afnasovo culture thus would have been forced first to veer west where they may have come into contact with Andronovans (Indo-Iranians) on their path to the south.

The model presented here is the most efficient in terms of geography, linguistics and archaeology (the evidence of physical anthropology remains moot). It requires the least distance for the Tocharians to journey into their historical seats and it provides them with a staging area which is, on the one hand, distant from the rest of the steppes to account for the independent evolution of a Tocharian stock of languages while also permitting them to come into contact with Indo-Iranians during their move south into East Central Asia. It also places them in that part of East Central Asia which is most closely associated with the steppelands to the north. If we suggest an alternative and attempt to pin Tocharian origins on a west-east migration across the Pamirs into East Central Asia, we will find it difficult to explain why it is the Saka who predominated in the western part of the Tarim. To argue that the Tocharians were the earliest colonizers from the west who were subsequently pushed eastwards by Iranians such as the Saka would leave the early Proto-Europoid peripheral Qawrigubh physical type in what would become later the Tocharian-speaking territories entirely anonymous. Moreover, there is no evidence in the Saka vocabulary that they absorbed the earlier vocabulary of already settled Tocharian farmers.

The European Card: The Long Trek

While the Afnasovo model may have much to recommend it, there are those who would reject its underlying linguistic premise: that the Tocharian language stock was somehow peripheral to the other Indo-European stocks. As we have seen, there are a considerable number of linguists who would regard Tocharian as a ‘western’ Indo-European language and insist that its origins cannot be on the outer edge of the Indo-European continent of dialects but rather that it must start farther to the west. Specifically, this means beginning the Proto-Tocharian movement from somewhere west of the Indo-Iranians, e.g. in eastern Europe (northwest of the Black Sea). Support for this model might be found in some of the textile evidence that relates the use of twills and plaids among the populations of East Central Asia with those of West Central Europe. Such a model then requires the ancestors of the Tocharians to move through the territories in which the Indo-Iranians were dispersing in order to arrive, again from the north, in the Tarim Basin. How robust is this model?

From an archaeological standpoint this renders the Proto-Tocharians archaeologically invisible until they emerge in the Tarim Basin. By this we mean that 1) there is no archaeological culture that makes a west to east trajectory during the 2nd or 1st millennium BC that we could conveniently associate with a Tocharian migration and 2) if the Proto-Tocharians are to be associated with one of the existing archaeological cultures, e.g. a phase or regional grouping of the Andronovo or other steppe culture, then they are indistinguishable from archaeological cultures to which we would assign an Indo-Iranian linguistic identity. To this it might be added that the basic trajectory of the steppe populations from at least the 1st millennium BC onwards was from east to west, carrying Cimmerians and Scythians into contact with farmers northwest of the Black Sea; a western origin for the Tocharians requires them to move upstream of the current of population movement. The only way this objection could be mitigated is by proposing a long trek to the east during a narrow window of opportunity in the centuries just before the general westward movement of steppe populations that began around the 8th century BC. Such a migration would have had to be quick and clean – so clean that it left no deep archaeological traces such as burials, settlements, etc.; we may find parallels between the textiles of Central Europe and the Tarim Basin but, so far, these do not include flying carpets.

There are linguistic problems with this model as well. A western origin for the Tocharians requires them historically to move across the face of prehistoric Indo-Iranians and to do so over a protracted period in their move towards the east. If such a movement had actually occurred, we might expect a far greater impact of Indo-Iranian on Tocharian than we find and those cultural loanwords which we do encounter borrowed from Indo-Iranian into Tocharian, i.e. words associated with irrigation agriculture, would be inexplicable.

The model of an origin of the Tocharians to the west of the Indo-Iranians, whatever its linguistic merits, is by far the weaker of the two models proposed. It is not that it can be shown to be impossible – archaeological evidence cannot be used in such a way – but it is archaeologically undetectable.

Conclusions

We may now offer as a working hypothesis the following model of ethnolinguistic development in East Central Asia.

1. The earliest Bronze Age settlers of the Tarim and Turpan basins originated from the steppelands and highlands immediately north of East Central Asia.
2. These colonists were related to the Afnasovo culture which exploited both open steppelands and upland environments employing a mixed agricultural economy.
3. The Afnasovo culture formed the eastern linguistic periphery of the Indo-European continuum of languages whose centre of expansion lay much farther to the west, north of the Black and Caspian seas. This periphery was ancestral to the historical Tocharian languages.
4. By about 2000 BC the Afanasevo culture, which was at that time being absorbed by the Andronovo culture from its west and other cultures in the Yenisei region, pushed southwards and came into contact with settled Indo-Iranians to the northwest of the Tarim Basin. Here they gained both the rudiments of irrigation agriculture and some of the Indo-Iranian terminology associated with it before they entered the Turpan and Tarim basins as the Proto-Tocharians.

5. Many of the Bronze Age mummies preserved in the archaeological record of East Central Asia may be assigned a probable (Proto-)Tocharian identity.

6. The descendants of these earliest Bronze Age colonists occupied the northern and eastern portions of the Tarim Basin and survived in their oasis settlements to emerge later as the occupants of Kocha, Qarakshahar and Turpan, leaving a residual linguistic legacy in Kroran.

7. Subsequent movements from the steppelands carried other peoples into the Tarim Basin. Those who settled in oases occupied by the Tocharians or in their vicinity were linguistically absorbed by them, while those who maintained a nomadic social structure moved with their herds around the peripheries of the Tarim and Turpan basins to be recorded in Han documents as Yuechi, Wusun and other possibly nomadic peoples. These were at least in part Iranian-speaking populations, although remnants of and combinations with the initial colonizers by Tocharians may also have been part of these societies. It is entirely possible that the ancestors of some of the mummies derived from these later intrusions.

8. Throughout the 1st millennium BC other Iranian populations, historically ancestral to the Saka languages, entered the Tarim Basin from the west and ensured that the western and southwestern portions of the Silk Road were Iranian-speaking. This population maintained its mobility and secured the spread of Iranian loanwords throughout East Central Asia which were adopted by the earlier Tocharians. This can be seen most clearly in the vocabulary of commerce and warfare.

These movements of both Tocharians and Iranians into East Central Asia were not a mere footnote in the history of China but, as we are about to see, were part of a much wider picture involving the very foundations of the world's oldest surviving civilization.

CHAPTER ELEVEN

Legacy

In the early 17th century Viscount Saint Alban, Baron of Verulam, or, as we usually encounter him in the history of literature, philosophy and science, Francis Bacon, wrote that the three most significant inventions in the history of humankind were printing, gunpowder and the magnetic compass. For him, the origins of all three inventions were 'obscure and inglorious'; for us, their origins lay along with a host of other important discoveries, in China. In some cases, the Chinese were simply first with their inventions, taking chronological priority over developments in the West. For example, the Chinese were the first to utilize iron ploughs for clearing wet and heavy soils while people in the West independently manufactured their own iron ploughs somewhat later. In other instances the list of Chinese improvements in the technology of the plough, particularly the curved mould board that permitted a vastly more efficient turning of the soil, had a direct impact on the West. Brought from China in the 17th century by Dutch sailors and then conveyed to England where the Dutch were employed to drain the East Anglian fens and Somerset moors, the Chinese plough became the model design for the subsequent European and American ploughs. By a similar route, 18th-century Europeans began to replace their winnowing