

**THE MYSTERY OF TOCHARIANS AND TOCHARIAN-SPEAKING PEOPLES  
IN WESTERN CHINA. LINGUISTIC EVIDENCE FOR THE PRESENCE  
OF INDO-EUROPEAN PEOPLES IN WESTERN CHINA  
— ANALYSIS OF CURRENT ISSUES**

**I. ARCHAEOLOGICAL CONNECTION — THE DISCOVERY OF THE “TARIM MUMMIES”  
AND THEIR LINKS TO THE MYSTERY OF TOCHARIANS.**

The region of Chinese Turkestan (present day Xinjiang-Uygur Autonomous Region) is a vast land covering 1,664,900 km<sup>2</sup>. Despite its harsh desert climate, there are many areas (oasis, river valleys) suitable for agriculture. The conditions are especially good for animal husbandry, so many of the native peoples lead nomadic life. Since the beginning of the Bronze Age, important migration routes leading from West to East passed through Xinjiang. As a result, through most of its history this inhospitable region was incredibly diverse both ethnically and linguistically. The first European expeditions dating back to the beginning of the 20<sup>th</sup> century discovered texts written in 17 different languages (Mallory, 2008, 44). According to Doug Hitch at least 20 languages denoted by 23 writing systems can be attested in the first millennium CE alongside the earliest trade routes that will later form the famous Silk Road (Anderson, 2012, 5; Hitch, 2010, 1). The following article is an attempt to present and systematically organize various linguistic, palaeographical and lexicographical approaches, in an attempt to confirm the presence of Indo-European, or Indo-European-speaking peoples in Chinese Turkestan as early as 2<sup>nd</sup> millennium BC. Existing theories have been critically evaluated in search for any convincing evidence.

Since the end of the 19<sup>th</sup>, the area of the TaklaMakand Desert in the Xinjiang-Uygur Autonomous Region, and many ancient oasis-cities located in the Tarim Basin attracted much attention from western archaeologists and early adventurers-explorers. For nearly five decades between the 1880' and the beginning of 1930' much of the regional cultural heritage was already appropriated by various colonial institutions and museums in Great Britain, France, Russia, Germany and Japan. Turbulent decades of the second half of the 20<sup>th</sup> century effectively closed the region for western science, stopping any further exploration efforts. It was therefore all the more intriguing, when between 1978 and 1994 one of the leading Chinese archaeologists from the Xinjiang Institute of Archaeology in Ürümqi, Prof. Wang Binghua, and his Uygur colleagues discovered, and exca-

vated Bronze Age and Early Iron Age European-Caucasoid mummies at *Qizilchoqa* near *Qumul* (Hami, 哈密), and various other important places alongside the ancient Silk Road. In spite of its importance, this discovery had to wait for a new opening in Peoples Republic of China politics, before it could be officially recognized by the western scientific community. In 1994 Prof. Victor Mair from the University of Pennsylvania, began a long-term cooperation with Wang Binghua, which led to many new discoveries in the region in the early 2000s. Today, the significance of those discoveries cannot be overstated. They constituted the first physical proofs for the Indo-European presence as far east as present-day Xinjiang and not only provided an insight into early migrations of peoples throughout Eurasia, but also meant that many theories about the history of China, the ancient Silk Roads, have to be re-examined and rethought.

In literature and popular discourse, the archaeological findings in question are often referred to collectively as “Tarim Mummies”. Most of them came from large burials near the ancient cities of Krörän (*Loulan*, 楼兰), Kucha (*Qinci*, 龟兹), Chärchän (*Qiemo*, 且末) and area near Xiaohe Oasis (also known as: *Xiaohemudi*, 小河墓地 — Xiaohe cemetery. First explored in 1934 by Folke Bergman, a Swedish archaeologist, but the site’s location was lost, until the Xinjiang Archaeological Institute rediscovered it again in 2000). Standard dating techniques (including carbon dating) placed the origins of those mysterious discoveries in the period between roughly 2000-1800 BC and 300 CE, if those estimations are correct, it could prove that Indo-Europeans migrated to Xinjiang long before any traditional estimations. The mummified bodies have a clearly Caucasoid physical appearance; they are much taller than any of the contemporary mongoloid peoples found in the Tarim Basin, have elongated bodies, angular faces, recessed eyes, blond, red to deep brown hair (Tomezzoli, Kreutz, 2011, 69-70). In most cases, the bodies are extremely well preserved as natural desert mummies (not mummified through any kind of artificial process but due to the natural conditions). The extremely dry climate allowed hair, skin, and even skin tattoos to survive intact. The objects found in the burial sites, especially textiles (analysis of the weaving techniques constitutes an important set of evidence) and horse saddles are remarkably similar to early Scythian craftsmanship. This fact allowed prof. Mair to hypothesize about their possible place of origin on the Pontic-Caspian Steppe and Eastern Ukraine (Mair, Mallory, 2000, 144-150). Which is where the Proto-Indo-Europeans supposedly originated, and from where they began their migration<sup>1</sup> (Mallory, Adams, 2006, 96-104). Although the oldest mummies dating back as early as 1800

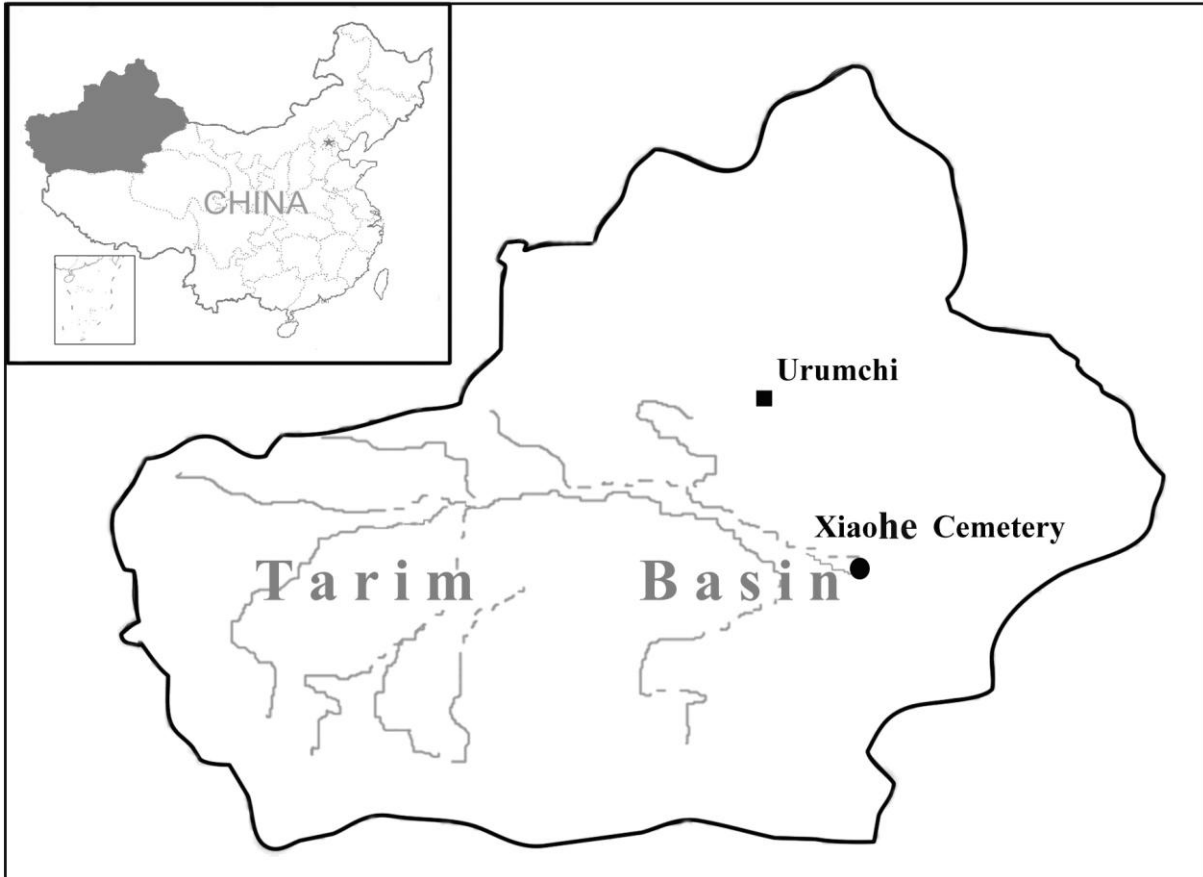
<sup>1</sup> So called „Kurgan theory” is most widely held, but not universally accepted hypothesis about the origins of Proto-Indo-Europeans, see: Gimbutas, Marija (1970), *Proto-Indo-European Culture: the Kurgan Culture during the Fifth, Fourth, and Third Millennia B.C.*, in Cardona, George; Hoenigswald, Henry M.; Senn, Alfred, *Indo-European and Indo-Europeans:*

BC were clearly Caucasoid, later discoveries dated to the period between 206 BCE — 9 CE showed an increasing number of mongoloid features, which clearly indicates that the population of TarimBasin had intense relations with neighbouring ethnic groups, including early Chinese states (Mair, 1995a, 3-4).

The “Tarim Mummies” could be the most important proof, that the earliest residents of the TarimBasin came from the West; likely as a conglomeration of nomadic, shamanistic communities (probably similar to a tribal federation; not unified politically, but related culturally and linguistically). Based on that assumption, Victor Mair put forward a following theory: the earliest mummies found in the TarimBasin are members of such an early, tribal community of Caucasoid or Europoid people, who arrived in the TarimBasin possibly as early as 3000 BC, likely through the Pamir Mountains. Their presence in the TarimBasin was strong for nearly 2000 years, and ended in stages, due to the arrival of the Chinese and other East-Asians following the Han Dynasty incursions into the central Asia (Mair, 1995a, 281-307). However, the exact time when the admixture of the East and the West occurred in this area is still a mystery. Probably such a process, due to its complexity, began much earlier before the first contacts with Chinese civilisation took place. There is a general consensus among experts that the last remnants of those peoples finally disappeared around the latter half of the 9<sup>th</sup> century CE, which coincided with the arrival of the Uygurs in 842 CE, after the collapse of the Orkon-Uygur Kingdom in Mongolia destroyed by the Kyrgyz (Tomezzoli, Kreutz, 2011, 69; Mair, 1995, 3-8). It seems obvious, that over two millennia of cultural presence in the region must have left more than only human remains and artefacts. The question of linguistic evidence that could be linked to those early western peoples remains a particularly interesting problem. The discovery of Xinjiang desert mummies could be interpreted as the crowning, physical evidence in a longstanding linguistic mystery, regarding a much earlier discovery of the enigmatic Tocharian language group.

---

*Papers Presented at the Third Indo-European Conference at the University of Pennsylvania*, Philadelphia: University of Pennsylvania Press, pp. 155–197



## II. TOCHARIAN LANGUAGE IN THE INDO-EUROPEAN CONTEXT

The Tocharian branch of the great Indo-European language family was the last one to have been identified, and in many respects still remains the least studied and poorly understood issue in the Indo-European studies at large. That the presence of a well-established language group complete with a variety of writing systems adopted to denote it, clearly indicates a well-developed civilisation. However for the better part of the 20<sup>th</sup> century the origins of the Tocharian civilisation in Central Asia were complete mystery, and intrigued a succession of prominent linguists including: Walter B. Henning, George S. Lane, Emil Sieg, Wolfgang Krause, Thomas Werner and Jia Xianli. the precise position of the Tocharian branch within the Indo-European language family remains a matter of much controversy. the classical approach to the problem linked Tocharian with Germanic languages, southern Slavic languages and some forms of Greek (Hamp, 1998, 307-346). A more modern approach is based on the assumption that Tocharian belongs to the first branches which separated from the Proto-Indo-European root, and together with Anatolian (traditionally viewed as the oldest) and similarly to Celtic languages in the West, Tocharian branch should be view in diachronic isolation form the others (Mallory, 2015, 32-35; Adams, 1989). There are some important linguistic arguments to substantiate that theory: firstly, Tocharian languages are very conservative in terms of phonetics, and possess a very complicated syntax

displaying many PIE characteristics. Secondly, Tocharian languages resemble western branches of the Indo-European language family (Hellenic, Celtic, Italic and Germanic.) in preserving the velar nature of PIE phonemes: /\*k/, /\*g/, /\*g<sup>h</sup>/. In Slavic, Indo-Aryan, Iranian and Baltic those phonemes changed into sibilants and fricatives (Blažek, Schawartz, 2008, 45; Mallory, Adams, 1997, 461). In Indo-European studies, it is generally accepted that an isogloss can be drawn, dividing the Indo-European languages into two geographical separated groups, so called: “centum” and “satem” (Szemerényi, 1990). Consequently, the first group represented by the western branches of the Indo-European language family, and the latter, languages located more to the east. Tocharian appears to defy that theory, being the eastern-most branch of the family, and still phonetically belonging to the centum group. This could indicate that Tocharian branch is indeed extremely old, and separated from the PIE and a very early time, before the changes occurred in the eastern Indo-European languages. Although it should be mentioned that today the centum-satem isogloss is often criticized, and many scholars doubt that such a division could be substantiated by evidence (Mallory; Adams, 1997, 461-462). The geographical isolation on the western-fringes of China helped to preserve Tocharian languages in their archaic form. It should be noted that Tocharian languages share remarkably little similarities with both the Iranian, and Indo-Aryan branches, despite belonging to the same language family, and their geographical proximity (Pinault, 2002, 244). Linguistically speaking the Tarim Basin is encircled and isolated by Iranian and Indo-Aryan languages to the south, Sino-Tibetan languages to the East and Altaic languages to the north. This means that most likely Tocharian speakers had to cross central Asia and arrive in Xinjiang before the region was dominated by Iranian and Indo-Aryan speakers (Mallory, 2015, 32-35). However since it is impossible to determine the precise time at which the Indo-European language family began to differentiate itself, therefore based only on diachronic linguistic evidence the precise age of the Tocharian ethnic group cannot be confirmed. Another curious feature of the Tocharian languages, potentially helpful in determining the origin of the ancient Tocharians, is that they have lost some of their case endings and replaced them with certain agglutinative traits. This is a highly unusual process for inflected languages (and Indo-European languages in general), and could suggest the involvement of a non-Indo-European influence in the grammatical evolution of Tocharian. Furthermore, since both Uralic and Altaic languages belong to the agglutinative type, some scholars suggested that the Tocharians could have entered the Tarim Basin by migrating alongside the northern part of the Eurasian Steppe and having some contacts with Uralic or Altaic speaking populations along the way (Mallory, 2015, 32-35). This however is merely a hypothesis and there are numerous arguments against it, for example: many other languages alongside different potential migration routes of the early

Tocharian population also display agglutinative properties, and furthermore none of the process involved in shading the cases in question requires an agglutinative component (the changes might have occurred naturally, which is unusual but not impossible).

The first Tocharian document appeared in Europe in the year 1892, it was a single page, published by Sergey Oldenburg who received it from a Russian consul in Kashgar. The text written in early form of Brahmi script represented an example of previously unknown language at that time regarded as a possible early Prakrit dialect (Blažek, Schawartz, 2008, 47). The document was identified in 1893, as one of the Buddhist sutras, by Leumann when based his research on the existing Sanskrit name references. In the following years, many other texts were discovered, some of them much older than Oldenburg manuscript. However they remained either unidentified, or undeciphered, and were rarely studied. Until finally, in 1908 German linguist Emil Sieg, with the help of his student Wilhelm Siegling, published a famous article: „Tocharisch, die Sprache der Indoskythen, Vorläufige Bemerkungen über eine bisher unbekannte Literatursprache“ in Berlin. Based on the existing bilingual materials they identified the language as previously unknown member of the Indo-European language family. Initially the name “Tocharian language” attributed to it was a subject of some controversy. It was first proposed by F. W. K. Müller (1907: 960), who based it on information from an Old Uyghur copy of a rare Buddhist text *Maitrisimit* (toch A: *Maitreyasamiti-Nāṭaka*), according to his research there was a clear indication that the text was translated from *tobri* (OUgr.) language. Müller’s proposition was accepted (not without some hesitations) by Emil-Sieg and Wilhelm Siegling, at that time leading experts in Tocharian studies. In 1916 Müller and Sieg proved that Old Uyghur *Maitrisimit* was a direct translation from *Maitreyasamitināṭaka* a Buddhist text written in Tocharian (Tocharian A). Sieg and Müller, linked this name phonetically with a well-known ethnonym *Tócharoi* (Ancient Greek: *Τόχαιοι*, Ptolemy VI, 11, 6, 2<sup>nd</sup> century AD), and proposed the name “Tocharian” (German: *Tocharisch*) (Beckwith 2009, 380–383). Those findings were further substantiated by a postulated link between the Old Uyghur word *tobri* and the Sogdian compound “*ctβ’r twyr’ystn*”; “four *Tuyric* countries” identified in the *Karabalgasun* inscription (line 19)<sup>2</sup>, which seems to refer to the area around *Qarašabr* (Yoshida, 2011, 530-533).

For some time after their initial identification, Tocharian texts were extremely rare, and in most cases incomplete. It seemed that Tocharian might share the fate of the Minoan language, and remain a complete enigma. Fortunately a crucial discovery was made in 1900, when Taoist monk Wang Yuanlu found the hidden library in Dunhuang (敦煌) temple complex in

<sup>2</sup> the trilingual inscription at Karabalgasun, in Old Turkic, Sogdian, and Chinese (F. W. K. Müller, “Ein iranisches Sprachdenkmal aus der nördlichen Mongolei,” *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, Phil.-hist. Klasse, 1909: 726-30.)

the present-day Gansu province of China. the discoveries of the so called “Mogao Caves” (莫高窟) draw the attention of many European orientalists, and in 1907 an expedition led by Sir Aurel Stein, a Hungarian-British archaeologist and indologist arrived at Dunhuang. Aurel Stein acquired more than seven thousand manuscripts including approximately one thousand two hundred in Tocharian (Wang & Perkins 2008, 8). Later, in 1908 a French sinologist Paul Pelliot examined the Mogao Caves and also brought many thousands of items back to France. Later the same year there was a Japanese expedition led by Otani Kozui. Finally in 1914, Sergey Oldenburg organized a Russian expedition. the remaining small number of manuscripts was collected by a Chinese scholar Luo Zhenyu and deposited in the recently established National Library of China (Hopkirk, 2006, 26-27; Stein, 1980). In total the Dunhuang library contained over four thousand texts written (partially or completely) in Tocharian, most of them, dating back to the period between 6<sup>th</sup> and 7<sup>th</sup> centuries CE (Wang, Perkins, 2008, 8).

During the next several decades many other texts were discovered in repositories and graves all over the Tarim Basin (the total number remains unknown). Particularly important were the findings from the former oasis-city states: Krörän (*Loulan*, 楼兰), Turpan (*Tulufan*, 吐鲁番) and Kucha (*Qiuci*, 龟兹). Based on the analysis of this vast corpus of texts, two distinct Tocharian languages were identified: Tocharian A (mainly used in Turpan, hence: Turpanian) and Tocharian B (mainly used in Kucha, hence: Kuchean), both closely related and mutually intelligible. Today, around 500 texts in language A, and more than 3000 in language B have been discovered (Adams, 2006, 382-383). This disproportion is usually attributed to the different status of both languages. It is suspected that Tocharian A, much older and more conservative served as a prestigious language of religion and Buddhist scriptures (possibly dead at that time), while more modern Tocharian B, was treated as the vernacular. Many of the texts written in Tocharian B are secular and contain knowledge about medicine, history and geography. Recently some scholars begun to speculate about a possible existence of a third Tocharian language designated as Tocharian C. the evidence supporting that claim are not conclusive, and mainly based on the analysis of proper names existing as borrowings in Prakrit and Sanskrit administrative texts discovered in Krörän. Though unconfirmed, Tocharian C hypothesis is particularly interesting, because the language appears to be much older than Tocharian A. Furthermore, most of the postulated Tocharian C materials have been discovered precisely in the areas where the oldest “Tarim mummies” have been found (Mallory, 2011, 50-53).

### III. ANALYSIS OF THE EXISTING CHINESE SOURCES CONTEMPORARY WITH THE TOCHARIAN LANGUAGE MATERIALS DISCOVERED AT DUNHUANG

From a purely linguistic point of view, Tocharians must have arrived in the present day Xinjiang before the beginning of the Common Era, because of the presence of Tocharian lexemes in the ancient Indian Prakrit texts dating back to the 2<sup>nd</sup> century BCE. If that assumption is correct, than Tocharian lexicon should have left some traces in the historical texts of the Chinese civilisation as well. From the earliest times the Chinese were very meticulous record keepers. Historical chronicles, annals and lexicons are one of the most important elements of Chinese literary tradition. In fact Chinese scholars devoted much of their time to cataloguing and recording history. Every dynasty during the imperial period had at least one official and numerous unofficial chronicles. Unfortunately substantial differences between Indo-European and Sinitic language families preclude the possibility of a direct transference of lexemes from one language system to another. Any existing acquisitions and borrowings would most likely be phonetically changed to a degree when it would be extremely difficult to identify them without any doubts. To complicate things even further, our knowledge of ancient Chinese phonology is very poor, and most reconstructions to date are incomplete. It is therefore important, to focus on a corpus of texts representing later, relatively well known and attested forms of Chinese, and to examine only certain groups of lexemes, most likely adopted directly from other language systems, namely: toponyms and etnonyms. With those methodological restrictions in mind, it stands to reason that the analysis of available Chinese sources must produce some answers.

The period between 6<sup>th</sup> and 7<sup>th</sup> CE, was a crucial time, when the Northern Silk Road was the most important trade route linking the Eastern and Central Asia and the West. Tarim Basin was divided between many independent city-states. the political situation was relatively stable, trade-driven economy flourished and the dominant religion was Buddhism. Tocharian language was widely used and was one of the most important tools in trade, diplomacy, as well as Buddhist religious practices. Most of the known Tocharian texts are dated to that period, and fortunately, there is a Chinese source that gives a detailed description of the political and cultural situation of the Tarim Basin at that time. In 629 Xuanzang (玄奘) a Chinese Buddhist monk, traveller and translator of Buddhist literature, left Chang'an the capital of Tang dynasty, and went to India. He compiled an account of his seventeen-year long journey in single work: *Datang Xiyuji* (大唐西域記) usually translated as "The Great Tang Records of the Western Regions". In the first section Xuanzang mentions the state of *Aqini* (阿耆尼):



阿耆尼國，東西六百餘里，南北四百餘里。國大都城周六七里。四面據山，道險易守。泉流交帶，引水為田。土宜糜、黍、宿麥、香棗、蒲萄、梨、柰諸菓。氣序和暢，風俗質直。文字取則印度，微有繪絹。服飾氈褐，斷髮無巾。貨用金錢、銀錢，小銅錢。王，其國人也，勇而寡略，好自稱伐，國無綱紀，法不整肅。(http://www.cbe.ta.org, 大正新脩大正藏經 Vol. 51, No. 2087, 大唐西域記)

*The state of Aqini is 600 li from east to west, and 400 li from north to south. the Capital is surrounded by walls 6 to 7 li long. There are hills at all sides. the roads are safe and easy to defend. There are numerous streams connected by channels to irrigate the fields. the soil is good for cultivating: millet, winter wheat, dates, grapes, plums, pears and other kinds of fruits. the air is delicate and pleasant and the customs of the people are honest. the written characters are almost likethat of India. They wear brown wool and cotton dresses and do not wear hats. For currency they use gold coins, silver coins and small copper coins. The king is native to the country, he is brave but pays little attention to plans and enjoys talking about his campaigns, the country has no annals and the laws are not strict.*

This rather laconic description provides certain clues as to the possible Tocharian heritage of the state of *Aqini*. the first and most important seems to be the name itself. A Sanskrit term *Agni* (अग्नि) refers to the users of Tocharian A and appears in most Indian sources of that time. In Sanskrit texts there are also two frequently used derivatives: */Agnideśa/* and */Agniviṣaya/*, both referring to the land roughly corresponding with the territory of Xuanzang's *Aqini* state. the same toponym appears in one of the oldest known Tocharian B texts, discovered in 1907 and dating to the period between 624 and 646 CE. In Tocharian A there is similar name: */Ārsī/*, used mostly in reference to the land and language, for example: */ārsī-kāntu/*: Tocharian A for language (Blažek, Shwarz, 2008, 53-55). the phonetic changes visible in the abovementioned examples appear to be regular and could be the result of a natural evolution. It is however difficult to determine, whether either of this names could be treated as pure ethnonym. According to Karlgren's reconstruction<sup>3</sup> of the Middle-Chinese, the name should be pronounced as: */Ā-g'ji-nji/*. There is a striking correspondence with both Sanskrit, and Tocharian, therefore it is likely that the people of *Aqini* could indeed be the descendants of ancient Tocharians, or at least share the Tocharian cultural tradition. If geographical coordinates provided by Xuanzang's are correct, *Aqini* should cover the area around present day Turpan, in the vicinity of which a large collection of documents in Tocharian A was found. Another interesting detail is the writing system. We know that Tocharian languages were often denoted by a form of alphabetical writing closely re-

<sup>3</sup> See: Karlgren, B.; 1954, *Compendium of Phonetics in Ancient and Archaic Chinese*. [in:] Bulletin of the Museum of Far Eastern Antiquities. Vol. 26: 211–367.

lated to the Brahmi script. In many cases, Tocharian texts were indeed written in Brahmi alphabet (so called “slander Brahmi”), that was most likely the case at the time of Xuanzang’s visit. In the neighbouring countries a further 20 different writing systems could be attested (often interchangeable between different language systems). Therefore, using a particular writing system must have been conditioned in some way; by cultural or historical factors. In this case, a probable cause could be an Indo-European language (Tocharian?).

The evidence that *Aqini* was a Tocharian state seems to be very convincing, but in *Datang-Xiyuji* there is a description of yet another city-state, which provides similar conclusions and reinforces that notion. Not far to the west from *Aqini*, Xuanzang travelled through the state of *Qiuqzhi* (屈支) also known as: *Qiuqi* (龜茲, simplified: 龟兹):

屈支國，東西千餘里，南北六百餘里。國大都城周十七八里，宜糜、麥，有粳稻，出蒲萄、石榴，多梨、柰、桃、杏。土產黃金、銅、鐵、鉛、錫。氣序和，風俗質。文字取則印度，粗有改變。管絃伎樂，特善諸國。服飾錦褐，斷髮巾帽。貨用金錢，銀錢、小銅錢。王，屈支種也，智謀寡昧，迫於強臣。(http://www.cbeta.org, 大正新脩大正藏經 Vol. 51, No. 2087, 大唐西域記)

*The state of Qiuqzhi has a 1000 li from east to west and 600 li from north to south. the state’s capital is surrounded by wall 17 li long. the soil is suitable for growing: wheat, rice, grapes, pomegranates, many kinds of pears, nairfruits<sup>4</sup>, peaches and apricots. the ground is rich in minerals: gold, copper, iron, lead and tin. the air is pleasant and the customs of the people are honest. the written characters are almost like that of India. They are better than (the people of) other countries in music (plying the instruments) and dancing. They cloth themselves in brocades, and wear head bands. For currency they use gold coins, silver coins and small copper coins. the King is native to the country; he has little wisdom and is ruled by powerful ministers.*

This description is very similar to that of *Aqini*. Apparently both countries shared similar customs and political systems. In both cases, there are writing systems of Indian provenance in use. However, the question of the name *Qiuqzhi* remains problematic. the kingdom was also known in old Uygur as /*Kūsān*/ which according to some scholars corresponds with Tocharian B adjective: /*kuśiñne*/ and noun /*Kuśi*/, both words were used in reference to the Kings and the state of Kucha (Blažek, Shwarz, 2008, 53-55). Therefore Chinese name *Qiuqzhi* could be interpreted as a direct phonetic rendering of the original Tocharian word. the correspondence between final affricates: Tocharian /*\*si*/ and Middle Chinese /-dzi/ appear plausible and could be expected. *Kucha* was one of the places where the biggest number of Tocharian manuscripts has been found. Most of the “Kuchean” manuscripts were written in Tocharian B language, using Tocharian al-

<sup>4</sup> 柰- a kind of apple

phabet. (Mallory, Mair, 2000, 270-296, 330-333). All the clues from *DatangXiyuJi*, could lead to a following assumption: both states — *Aqini* and *Qiuqzhi*, shared (at least to some extent) Tocharian culture and language. At the time of Xuanzang's visit they were both old countries with well-established cultural traditions. Based on this assumption, we can tell that, in the 7<sup>th</sup> century CE, at least two well-developed Tocharian states existed in the Tarim Basin. The possibility, that there were indeed Tocharian-speaking and (possibly) ethnic Tocharian countries in Xinjiang between 6<sup>th</sup> and 7<sup>th</sup> centuries CE seems to be very high. However, a pertinent question still remains: where did the ethnic Tocharians come from? In order to look for linguistic clues to answer that question, we must take a step back, and examine much older texts.

#### IV. ANALYSIS OF THE EARLIER CHINESE LANGUAGE SOURCES, IN THE CONTEXT OF THE EXISTING STUDIES AND HYPOTHESIS

As mentioned before, there are many theories about the origins of the Tocharian People. One (perhaps most influential) is the Mair hypothesis, based on the examination of “Tarim Mummies” and placing the origins of the Tocharians on the steppes of Eastern Europe (See subsection I). There is however another, less known theory, very interesting from the linguistic point of view. In 1978, W. B. Henning published an article in “Society and History” entitled: “The First Indo-Europeans in History”. In this article he claims, that most probably Tocharians originated in the Middle East. He identifies them with two ethnic groups mentioned in the cuneiform documents of Babylonia. In the Akkadian language they were known as *Guti* and *Tukri*. According to the Babylonian sources, the *Guti* people came from Western Persia. They waged war against Babylonian Kingdom under the king *Narâm-Sin* (2254–2218 BC), subsequently conquered and ruled the whole of Babylonia for almost 100 years: c. 2000-2100 BC. *Tukri* territories bordered the *Guti* domain in the east or southeast. Later Henning speculated that both the *Guti* and the *Tukri* left Persia before the end of the 3<sup>rd</sup> millennium BC. They arrived in what is now Western China, later some of their original tribes settled in permanent habitations; others clung to the nomadic life. This theory appears logical since all the dates match, and even the proposed time of migration matches the earliest “Tarim mummies” c. 1800 B.C. (Yu, 2010, 44-45; Henning, 1978). Furthermore, in the oldest Chinese historical texts there are two frequently mentioned nations, which could resemble the *Guti* and *Tukri* — *Daxia* (大夏) and *Yuezhi* (月氏).

The name *Daxia* (大夏) is old, and appears in *Zuoqibuan* (左傳); in the section devoted to the first year of the Duke of Zhao of Lu (魯昭公), where it is mentioned as a vassal state to the Zhou dynasty, located in the west. It should be noted, that is the only reference to *Daxia* as

a state. In other sources like: *Lüshi Chunqiu* (呂氏春秋), *Shiji* (史記), and *Guanzi* (管子) *Daxia* appears mainly as a toponym (often as enigmatic “Ruins of *Xia*”), located between *Yicheng* (翼城) and *Taiyuan* (太原) (Yu, 2010, 4-5) There is evidence that *Xia* people inhabited an area in the present day Gansu province (called: *Hexi*, 河西) and early before the end of Zhou period migrated to the present day Shanxi. In some of the Greek sources, mainly Geography of Ptolemy<sup>13</sup> (VI, 16) the *Thaguri* people, *Thaguri* Mountains and *Thogara* town appear to have been located in *Hexi* region. This has been further confirmed by studies of Tibetan and Khotanese documents. *Thaguri* and *Thogara* can be regarded as different transcriptions of toponym *Daxia*, or even be derivatives of the Greek name of Tocharians: *Tókbaroi* (Τόχαροι). This shows that the Tocharian people probably dwelled in the *Hexi* region as early as the Spring and Autumn Period (771-476 BC) (Yu, 2010, 8-9). In the *Shiji* (史記), section: *Dayuan Liezhuan* (大宛列傳) there is a mention that the state of *Daxia* in the Western Regions was located on the southern banks of the River *Gui* (媯) (Amu Darya). Later in c. 130 BC the *Daxia* people were conquered by the *Da Yuezhi* (大月氏). Again, some scientists view the name *Dayuan* (大宛) as version of the original Tocharian name. The Middle Chinese form of that name was probably pronounced as /\**Taxwār*/\* (Pulleyblank, 1962, 90). This could be confirmed by later reflections such as toponyms: *Tubuoluo* (吐火羅), *Tubuluo* (吐呼羅), *Dubuoluo* (覩貨邏), or a modern Tarim Basin village name: *Tuobula* (托呼啦), located not far to the west from Khotan (Li, 2006, 15; Yu, 2010, 12).

Another ethnonym/ toponym that must be considered in context of Chinese historical sources: *Yuezhi* (月氏). In *Yi Zhou Shu* (逸周書), the name *Yuezhi* appears in the list of tributaries to the Zhou dynasty. It is mentioned that *Yuezhi* supplied horses for the Zhou army. From *Guanzi* (管子) a text covering period c. 5<sup>th</sup> – 1<sup>st</sup> BC, section: *Qingzhong Yi* (輕重乙), comes following information about *Yuezhi*:

(...) 玉出於禺氏之旁山，此皆距周七千八百餘里 (<http://ctext.org/guanzi/qing-zhong-ii>)

(...) *Jade comes from the hills near Yuzhi, it is (this place) located 7,800 li from Zhou.*

In this passage *Yuezhi*, appear as *Yuzhi* (禺氏) which is another frequently used name (possibly more archaic). They were using jade, which came from mountains at least 7, 800 li to the west from Zhou. Pulleyblank identifies that area as present day Chotan within the Tarim Basin (Pulleyblank, 1966, 19). In *Shiji* (史記), there are also references to several wars between the *Yuezhi* and *Xiongnu* (匈奴) a nomadic people from the steppes of central Asia, waged even before

the rule of the Qin dynasty. There is also some information about the *Yuezhi* dwelling between *Dunhuang* (敦煌) and *Qilian* (祁連):

始月氏居敦煌、祁連間，及為匈奴所敗，乃遠去，過宛，西擊大夏而臣之，遂都媯水北，為王庭。其餘小眾不能去者，保南山羌，號小月氏。(http://ctext.org/shiji/dawan-lie-zhuan#n9030)

*At the beginning the Yuezhi dwelled between the Qilin and Dunhuang, (later, after) they were defeated by the Xiongnu and they moved far to the west, beyond Dayuan. There they conquered the people of Daxia and establish the King's court on the northern bank of the River Gui. A small number of their people who were unable to make the journey west, found refuge among the Qiang of the Southern Mountains, where they are known as the Xiao Yuezhi.*

*Shiji* also informs us about a possible split that took place in the 3<sup>rd</sup> century BC; a part of the *Yuezhi* moved north to *Yanmen* (雁門) and replaced the *Sai* (塞) tribes that lived there. They managed to build a strong state and monopolized the trade between east and west for a time, in *Shiji* they are known as *Da Yuezhi* (大月氏). They were defeated by *Xiongnu* in 177 BC. Others who left the former land of *Yuezhi* and moved to the northwest from Gansu were known as *Xiao Yuezhi* (小月氏) (Yu, 2010, 55). In both *Shiji* and *Hanshu* (漢書) there is a note that in c. 108 BC *Xiao Yuezhi* subordinated to the rule of Han dynasty. In both sources the King of *Xiao Yuezhi* is known as *Ruoju* (若瓊) Based on the Middle Chinese reconstructed form: /nyák-tsjó/, Pulleyblank hypothesised that the name could be related to a Tocharian adjective /ñäkc(i)ye/ which translates as “divine”, “heavenly”. According to this theory the ruler of *Xiao Yuezhi* was titled as “heavenly” or “celestial” king (Pulleyblank, 1966, 19-20; Blažek, Shwarz, 2008, 54). Furthermore Baxter hypothesizes the first character of the *Yuezhi* name: 月, in Old Chinese was pronounced as: /\*ŋ<sup>w</sup>jat/ (Baxter, 1992, 806) and could have been pronounced in an archaic north-western dialect as /\*tokwar/ or /\*togwar/, a form that resembles the Bactrian name /Toxoaq/ (/Toxwar/ ~ /Tuxwar/) and the medieval form /Toxar/ ~ /Toxâr/. (Hitch, 2010, 654-658)

According to the Henning's theory, both *Gutian* and *Tukri* could have shared a single cultural and linguistic tradition, the fact later reflected in the existence of two Tocharian languages. But the amount of evidence to support the claim that both groups should be identified with *Daxia* and *Yuezhi* seems inconclusive. Certainly, over the years there have been many speculations, but Chinese historical references to *Daxia* and *Yuezhi* are very sparse. Of course the speculations could be far more extensive, and the above presented analysis of Chinese texts is very rudimentary. However it seems obvious, that any further analysis of texts such as *Shiji* or *Zuo Zhuan* should

be very careful. Unfortunately it is easy to sometimes lose the sense of logic in idle speculations based only on very short references, or imprecise geographical data.

The hypothesis can be at least partially substantiated when we examine the historical events from non-Chinese point of view. According to Chinese sources the *Yuezhi* never appear as under the name “Tocharians” until they moved west to the valley of the Amu Darya, when they were driven out by the *Xiongnu*. In Indian, Persian, Sogdian and Greek sources from that period they are known as: */Tukhāra/*, */Toyri/*, */Tókharoi/* and Bactria itself came to be known as */Taxāristān/* ‘land of the Tokharians (Yu, 2010, 45; Henning, 1978). It is tempting to see this change of names as sudden, and unexplained. Seemingly the *Yuezhi* left China and arrived in Bactria as Tocharians. However *Toxāristān* has a corresponding name in the Chinese historical books, that is: *Daxia*. As it was mentioned above, the *Yuezhi* established their state in the valley of the Amu Darya River after they conquered the state of *Daxia*. To be sure, the *Yuezhi* are different from the *Daxia*. In other words, the *Guti* and the *Tukri* went their separate ways for a long time. Purely from a theoretical point of view the *Yuezhi* might indeed be related with the historical Tocharians of Afghanistan, but their history is fairly well attested in various Chinese and non-Chinese sources, and there is no direct evidence that they spoke Tocharian or that there is any linguistic connection between the speakers and writers of Tocharian in the Tarim Basin and the nomadic *Yuezhi* spread across vast distances in Central Asia (Mallory, 1989, 60).

## V. THE ISSUE OF COMPARATIVE LINGUISTIC RESEARCH RELEVANT TO THE SUBJECT

Analysis of the historical texts alone cannot provide sufficient information for claiming that the mysteries of the identity of the ancient Tocharians have been solved. It is interesting, but the available material is too small to be conclusive. Therefore as a final attempt we must examine the purely linguistic perspective. The evidence that there were some contacts between Tocharians or Tocharian speaking peoples and the Chinese is very strong. There are some traces of alien influence in the names and titles of ancient Chinese, and the onomastic analysis of their history proves that they were most likely Tocharian. So are there any other borrowings in Chinese, which could point us in the right direction? This type of diachronic analysis is tempting, but extremely difficult for several reasons:

Firstly, known Tocharian texts are relatively late and large percentage of them belongs to the Buddhist tradition. This fact greatly narrows the vocabulary, since only a small number of attested words will pertain to everyday life. Secondly, the available material contains great number of loan words from Sanskrit, Prakrit and Iranian, so the core lexicon is very limited and very likely contaminated. Thirdly, due to the relative lateness of the textual materials (VI-VIII CE),

Proto-Tocharian reconstructions, cannot reach beyond the 4<sup>th</sup> century BC. This means that there is a considerable gap between the period of possible Tocharian contacts with the Chinese and the reconstructed proto-Tocharian forms, and therefore the analysis is of questionable credibility. Finally, comparative- historical analysis between Indo-European language and Chinese must be very cautious. There are many methodological dangers on the way. the most important are the basic differences between the phonetics of both languages. Most of the Ancient and Middle Chinese words are monosyllabic; the number of possible syllabic combinations is quite small. Due to this simple fact it is relatively easy to find some correspondence between the words of Indo-European language and Chinese, especially if the phonetic-semantic correspondence is treated lightly. It is therefore crucial, to establish a fundamental rule: during the course of the analysis, both Tocharian and Old Chinese words must match semantically and phonetically. Furthermore, in order to deal with the abovementioned problem of the diachronic reconstructions, a broader reference material is needed, so the analysis includes reconstructed PIE forms if such exist. (Lubotsky, 1998, 380-382)

According to this there are only few words that could be considered possible Tocharian or Indo-European loan words in ancient Chinese. Most scholars suggest that the group should primarily include words describing elements of a chariot, as well as several words describing cities and fortifications. There is much archaeological and historical evidence that both chariots and strongholds came to China from the west. Three frequently analysed examples are:

Chin. 乘 *sheng* (chariot, with four horses)' < EC /zyingll/ < OC <sup>^</sup>/LjmgS/\* /Längs/, Toch.B /klenbe/, 𨋖 /klank/ (vehicle), Skt. /yāna/-, /vāhana/-, Toch. AB /blänk/- (to ride, travel by vehicle), PIE /\*kleng/.

Chin. 垣 *yuan* (wall), 園 (garden) < EC /peli/ < OC /<sup>^</sup>pek/, /\*peh/ ,Toch. B /pkante/, Toch.A /pkaiit ' / (obstacle) < PToch. /<sup>^</sup>pabnte/

Chin. 城 *cheng* (city wall, fortified wall) < EC /dʒ<sup>h</sup>eng/ < OC /\*djeng/, /\*deng/ < Toch. AB /tank/ < PIE /\*teng<sup>h</sup>/. (Lubotsky, 1998, 385-387)

Inter-linguistic comparison demonstrates little interaction between Old Chinese and Tocharian, although increasing word exchange with Indo-Iranian languages is documented during the time of the Silk Road. Furthermore, Tocharian itself shows little relation to its geographic neighbours. Indeed, its closest linguistic relatives are probably the Celtic, Italic, and Anatolian languages. the above reconstructions are very convenient from the Indo-European point of view, but in terms of Chinese, many of the examples do not fit perfectly. Aside of the phonetic inconsistencies often due to the insufficient knowledge about the evolution of Chinese, there are some important questions like for example in case of the last word *cheng* 城 closely connected to

the verb 成 *cheng*: “to achieve”, “to complete”. In this case there is no semantic relevance. Unfortunately many scholars still use the classic comparative method to relate Tocharian and Old Chinese. In many cases this approach produces misleading results.

#### SUMMARY:

Aside from all the above mentioned arguments, the hypothesis that Indo-European language came to be spoken by the *Yuezhi* is even more speculative. The Henning’s assertion that the *Guti* (*Yuezhi*?) and *Tukri* tribes spoke proto-Tocharian when they left Babylonia has absolutely no linguistic foundations beside the few coincidental references between some proper names. There are no texts, no physical evidence of any language that can be attested to either the *Yuezhi* or the *Guti* and *Tukri* peoples. Some scholars like Haskins took this argument a step further proposing that the *Yuezhi*/*Guti* were the *Massagetae* (*Μασσαγέται*) described by Herodotus; *Massa-getae* means “Great Getai (*Guti*?)” just as *Da Yuezhi* means “Great *Yuezhi*”, a people that he assigned to the Altai region of the eastern Eurasian Steppe. Besides this problematic use of linguistic and historical data, there are two immediate problems with the “*Yuezhi*–*Guti*–*Tocharian*” analogy. First, the languages association between texts from the documents excavated in the Tarim Basin and Tocharian language was arbitrarily made by Western scholars at the beginning of 20<sup>th</sup> century. This assumption was based on the Greek and Indic texts in which *Tocharoi*/*Tukhara* peoples are mentioned in the area of Tarim Basin (Mallory, Mair 2000, 333). This is no evidence at all, it is just an extrapolation of certain unrelated facts, fitting perfectly into place, assuming that one has a preconceived idea of what one wants to find. Secondly, there is no information in texts, nor any kind of artefacts, or other physical evidence found during any excavations in Xinjiang that could substantiate the link between any of the analysed names and the Tocharian languages. The names: *Yuezhi*, *Daxia*, *Guti*, and *Tocharoi* could only be ethnonyms placed upon non-literate tribes by their literate neighbours in a purely random way. Despite those reservations we have Chinese historical texts indicating that speakers of the Tocharian B language lived in the northern Tarim Basin some time before the invasion of the Turkic-speaking Uygur tribes in the eighth century CE (Mallory, Mair 2000, 280). Based on this fact, a further analysis of Chinese sources should be conducted. If physical evidence linking the “Tarim mummies” and any of the abovementioned peoples, could be found confirming one of them as Tocharians, than certainly it could only be verified through reference to historical sources (Thornton, Schuur, 2004, 91-92).

The final summary of this article must be inconclusive. This was merely an attempt to show the various directions in which our attempts to identify the Tocharian people could proceed. Based on the above analysis we can neither prove nor disprove any of the various hypotheses.



There are however two things that must be pointed out: for one, analysis of Chinese sources definitely produces interesting results. A much more thorough analysis, encompassing many more of the available texts should be conducted in order to find more patterns and cross-references in search for the traces of Tocharian history. Secondly, purely linguistic analysis, giving our current understanding of Tocharian language and the available lexicographical material cannot produce any credible results beyond proving few historical interactions (mainly onomastic). It is however not futile, Chinese texts contain many more names and words that must be examined but always according to strict methodology and in context of concrete historical evidence. Nevertheless any results produced using classic linguistic reconstruction techniques must be treated with a degree of caution.

The identification of the Tocharian people using only historical linguistic and comparative methodology is not possible. In order to finally produce any conclusive evidence, further anthropological, archaeological and genetic research is required. Only when all four approaches: historical and linguistic (based on the analysis of Chinese materials in reference to Indian, Persian and Greek sources), archaeological and genetic (based on analysis of the Tarim mummies) are coordinated final goal can be achieved.

#### Literatury:

- Adams D.Q.; 2006, *Some Implications of the Carbon-14 Dating of Tocharian Manuscripts*, Washington DC: Journal of Indo-European Studies, vol. 34
- Adams D.Q.; 1984, *the Position of Tocharian among the Other Indo-European Languages*, Ann Arbor: Journal of the American Oriental Society, vol. 103
- Anderson M.; 2012, *the Languages and Writing Systems of the Tarim Basin*, Philadelphia: Sino-Platonic Papers, vol. 228
- Balžek V., Schwarz M.; 2008, *Tocharians- Who they were, where they came from and where they lived*, Poznań: Lingua Posnaniensis
- Baxter W.; 1992, *the handbook of Old Chinese*, Berlin-New York: De Gruyter
- Beckwith Ch. I.; 2009, *Empires of the Silk Road a History of Central Eurasia from the Bronze Age to the Present*, Princeton: Princeton University Press
- Beckwith Ch.; 2009, *Empires of the Silk Road: A History of Central Eurasia from the Bronze Age to the Present*, Princeton: Princeton University Press
- Müller F. W. K.; 1907, *Beitrag zur genaueren Bestimmung der unbekanntenen Sprachen Mittelasiens*, [in:] *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften*, Berlin: Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften
- Hamp E. P.; 1998, *Who were the Tocharians?: Linguistic subgrouping and diagnostic idiosyncrasy* [in:] Mair V. H., [ed.] *the Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, Philadelphia: University of Pennsylvania Museum Press
- Henning W. B.; 1938, *Argi and the Tokharians*, London: Bulletin of the School of Oriental Studies, vol. 9, no. 3

- Henning W. B.; 1978, the *First Indo-Europeans in History*, [in:] Ulmen G. [ed.], *Society and History: Essays in Honor of Karl August Wittfogel*, Hague: Mouton
- Hitch D.;2010, *Empires of the Silk Road: A History of Central Eurasia From the Bronze Age to Present*, Ann Arbor: Journal of the American Oriental Society
- Hopkirk P.;2006, the *Lost Treasures of Central Asia*, Oxford: Oxford University Press
- Karlgren, B.; 1954, *Compendium of Phonetics in Ancient and Archaic Chinese*, Stockholm: Bulletin of the Museum of Far Eastern Antiquities,vol. 26
- Li Y.;2006, *HetianChunqiu (和田春秋)*, Ürümqi:Xinjiang RenminChubanshe(新疆人民出版社)
- Lubotsky A.;1998, *Tocharian loan words in Old Chinese: chariots, chariot gear, and town building*, [w:] Mair, V.H. [ed.] *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*. Philadelphia: University of Pennsylvania Press
- Mair V. H.;1995a,*Prehistoric Caucasoid Corpses of theTarimBasin*.Washington DC: Journal of Indo European Studies, vol.23
- Mair V. H., Mallory J.P.;2000,*TheTarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*,London:Thames & Hudson
- Mair V. H.;1995, the *Mummified Remains Found in the TarimBasin*. Washington DC: Special Issue of the Journal of Indo-European Studies
- Mallory J. P.;2011, *Bronze Age Languages of TheTarimBasin*, Expedition vol. 52, no. 3, Philadelphia: University of Pennsylvania Museum of Archeology
- Mallory J. P.;2015,*The Problem of Tocharian Origins: An Archaeological Perspective*, Philadelphia: Sino-Platonic Papers, vol. 259
- Mallory J.P., Adams D. Q.;2006, the *Oxford Introduction to the Proto-Indo-European and the Proto-Indo-European World*, Oxford: Oxford University Press
- Mallory J.P.; 1989, *In Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames & Hudson
- Mallory J.P.; Adams, D.Q., [ed.]; 1997, *Proto-Indo-European*.[in:] *Encyclopedia of Indo-European Culture*,Chicago: Fitzroy Dearborn Publishers
- Pinault G.J.; 2002,*Tocharian and Indo-Iranian: Relations between two linguistic areas, in Indo-Iranian Languages and Peoples* [ed.] N. Sims-Williams,*Indo-Iranian Languages and Peoples*, Oxford: the British Academy Press
- PulleyblankE. G.; 1966, *Chinese and Indo-Europeans*, London: Journal of the Royal Asiatic Society
- Stein A.; 1980, *Serindia: Detailed report of explorations in Central Asia and westernmost China*, vol. 5, London & Oxford: Clarendon Press
- Szemerényi O. J. L.;1990, *Introduction to Indo-European Linguistics*.Oxford: Oxford University Press,
- Thornton Ch. P.,Schurr T. G.;2004, *Genes, Language and Culture: an Example from the TarimBasin*, Oxford: Oxford Journal of Archeology, vol. 23, no. 1
- Tomezzoli G, Kreutz J.;2011, the *linguistic position of the Tocharian*, Proceedings of the 9<sup>th</sup>international topical conference –The Origin of Europeans, Ljubljana:Jutro
- Wang H., Perkins J.,[ed.];2008, *Handbook to the collections of Sir Aurel Stein*, London: the British Museum Occasional,vol. 129
- Yoshida Y.;2011, *KarabalgasunII. the inscription*, [in:]EncyclopædiaIranica XV, New York: Routledge&Kegan Paul
- Yu T.;2010, *TheEarliestTocharians in China*, Philadelphia: Sino-Platonic Papers, vol. 204.

**Internet sources:**

<http://www.cbeta.org>, 大正新脩大正藏經 Vol. 51, No. 2087, 大唐西域記. [Accessed]: 14.07.2018

<http://ctext.org/guanzi/qing-zhong-ii>. [Accessed]: 14.07.2018

<http://ctext.org/shiji/da-wan-lie-zhuan#n9030>. [Accessed]: 14.07.2018