

Significant (Z|-4|) admixture signal with a source from ancient Wusun observed in contemporary Bulgarians.

By Svetoslav Stamov, MS, Duke University, Association Science
Bulgaria

Abstract

We report the presence of significant Central Asian ancestry in both contemporary Bulgarians and in early medieval population from SMC (Saltovo Mayaky Culture) .

The existence of Chalcolithic-Iran (Hajj-Fruz) and Wusun related ancestral component in contemporary Bulgarians comes as a surprise and sheds light on both migration route and ethnic origins of Proto-Bulgarians. We interpret these results as an evidence for a Central –Asian connection for the tribes, constituting the population of SMC and Kubrat’s Old Great Bulgaria in Pontic steppe from 6th-7th century AD.

We identify Central Asian Wusun tribes as carriers of this component on the base from the results from f3 and f4 statistics. We suggest that Wusun-related tribes must have played role (or might have even been the backbone) in what became known as the Hunnic migration to Europe during 3rd-5th century AD. Same population must have taken part in the formation of the SMC (Saltovo-

Mayaki Culture) and Great Old Bulgarian during 6th-9th century AD in Pontic – Caspian steppe.

We also explore the genomic origins of Thracians and their relations to contemporary Europeans. We conclude that contemporary Bulgarians do not harbor Thracian-specific ancestry, since ancient Thracian samples share more SNPs with contemporary Greeks and even contemporary Icelanders than with contemporary Bulgarians.

Dead bones and broken pots do not speak languages, which in the absence of written testimonies creates a challenge in identification and interpretation of the historical processes that led to the genesis of contemporary Bulgarian ethnicity. While the language of the Proto-Bulgarians has been identified as belonging to Altaic language family, the scarcity of written resources and the inconclusive archeological evidence from the times of First Bulgarian Empire have created room for two-century long scholarly debates on the location of Proto-Bulgarian homeland that have not been set conclusively. (Rashev, 2005).

Genome-wide analysis of ancient DNA has transformed our understanding for past events and in the absence of written testimonies and clear archeological record (Haak, Reich et al, 2015) has proved to be useful addition to archaeological examinations and comparative linguistics.

Recent advancements in ancient DNA research have demonstrated that while pots are not people, people are not trees – Western Eurasia has been shaped by numerous folk migrations as migrations did play sometimes-decisive role in the peopling of Europe and to the historical processes that gave shape to it. Thanks to the discoveries coming from the field of archeogenetics, archeologists and historians finally stepped on secure evidence that migrations did occur. Once the facts of migrations have been conclusively established, archeologists and historians can finally move on to the next step. (Kristiansen 2017, D. Reich 2015, Mattieson 2018).

During the last 10 years, various interdisciplinary teams consisting of historians, archaeologists, population geneticists, linguists and anthropologists sequenced more than 2400 ancient samples from Eurasia, for a time period spanning from 40 000 YBP to 2000 AD. The results from their work gave the archeological community a tool for a high-resolution view of Eurasian past. By comparing and mapping the extracted DNA from people living in Eurasia from different eras, scientific teams gave the audience map of the population movements in Eurasia for the last 10 000 years. Their work has been employed for further research by both historians and anthropologists in order to build clearer models for the events from past times that have been camera obscura due to absent material and linguistic evidence.

In this survey, we employed ADMIXTOOLS 2

(<https://github.com/uqrmaie1/admixtools>, publication pending) to analyze 1240K capture SNP data from contemporary Eurasian sample against genomic datasets from 2400 ancient individuals, which we downloaded from <https://reich.hms.harvard.edu/datasets> (David Reich Lab, Harvard University)

We analyzed all contemporary and ancient samples using supervised and unsupervised graphic modeling, f_2 , f_3 , and f_4 statistics.

Results

We first analyzed and modeled modern Bulgarians in the context of contemporary western Eurasian populations.

Contemporary Bulgarians are admixture of Slavs, Proto-Bulgarians and Latinized Balkan populations from late antiquity that came into being after the establishment of FBK (First Bulgarian Kingdom) in 7th century AD. In terms of modern European populations, present Bulgarians can be modeled as an admixture of contemporary Greeks and contemporary Lithuanians, as proved by f_3 -statistics in the form of qp3pop. (Fig1). Modern Bulgarians exhibit signal of admixture from both Lithuanians and Greeks, as shown by negative f_3 -statistics in the form of f_3 (Bulgarian;

Greek, X). We computed lowest Z-score for f_3 (Bulgarian; Greek, Lithuanian) = -0.00296 (Z=-8.5).

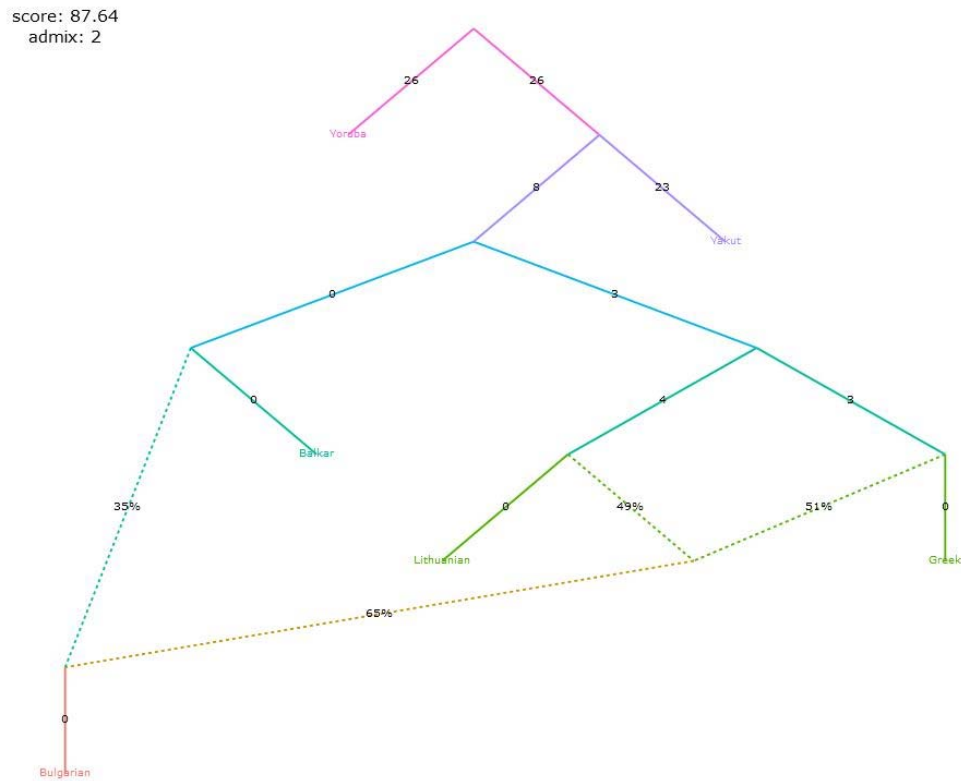
Fig. 1

| ADMIXTOOLS 2 | | | | | | | |
|--------------|-----------|-------|------------|---------|--------|---------|------------|
| | pop1 | pop2 | pop3 | est | se | z-score | p.value |
| | Bulgarian | Greek | Icelandic | 0,00 | 0,0021 | -1,44 | 0,14889776 |
| | Bulgarian | Greek | Lithuanian | -0.0055 | 0.0006 | -8.67 | 4.41e-18 |

There is no doubt that results reflect the settlement of Slavs and Proto-Bulgarians in SE Europe during late antiquity and their massive presence in FBK, which gave rise to contemporary Bulgarians.

We however established, that best models for modern Bulgarians require the addition of a third group next to Lithuanians and Greeks. We used Admixtools 2 function *qpGraph* to compute the admixture graph with and we established that best fitting model for contemporary Bulgarians via modern ethnicities should include a 35% admixture component coming from contemporary Caucasus populations . (fig.2).

Fig 2.



Since Caucasus-Iran related component has been presented in Europe from late Neolithic times and reinforced on multiple demographic events (Fernandes, D.M., Mittnik, A., Olalde, I. *et al.*, 2020), we investigated the relationship between contemporary Bulgarians and all publicly available ancient samples from

Caucasus, Central Asia and Iranian Plateau in order to identify the source of this admixture component.

In the beginning of our survey we put Bulgarian samples in the context of all ancient Eurasian samples we had available (2400 samples). We began with prehistoric populations that have been established to be the source of all contemporary European populations: West European Hunter-Gatherers (WHG), European Neolithic farmers (EEF) and Late Neolithic/ Early Bronze Age Yamnaya Pastoralists. We modeled modern Bulgarians as a mixture of WHG, EEF and Yamnaya. In order to achieve the most realistic results, we used the samples with the highest SNP coverage, with Loschbour sample for WHG (*Lazaridis, D. Reich et al., 2014,*) samples I0707 for Anatolian farmers (EEF) and Samara - I0231 , for Yamnaya. To test for differential relationship between target samples and ancestral populations, we later added two more Yamnaya samples (Ukraine and Caucasus) as well as second Loschbour sample to increase source variations. To clear the relationship within the Neolithic component in contemporary Bulgarians, we later added Germany _LBK sample and Chalcolithic Iran_Hajj Fruz sample. We used contemporary Yoruba and contemporary Yakut samples for out-groups as we built bigger out-group / reference group later. To avoid bias, we used only SNPs with no missing data among samples. We used ADMIXTOOLS 2 qpAdm *to estimate* the ancestral proportions coming from each source population.

We established that while half of Bulgarian ancestry could be derived from European Neolithic farmers ancestry, contemporary Bulgarian samples also harbor previously unreported, distinctive ancestry from Zagros Mountains in West Iran, exhibiting connection to Iranian chalcolithic sample Ch_Iran_HajjFruz (Fig. 3).

Fig 3.

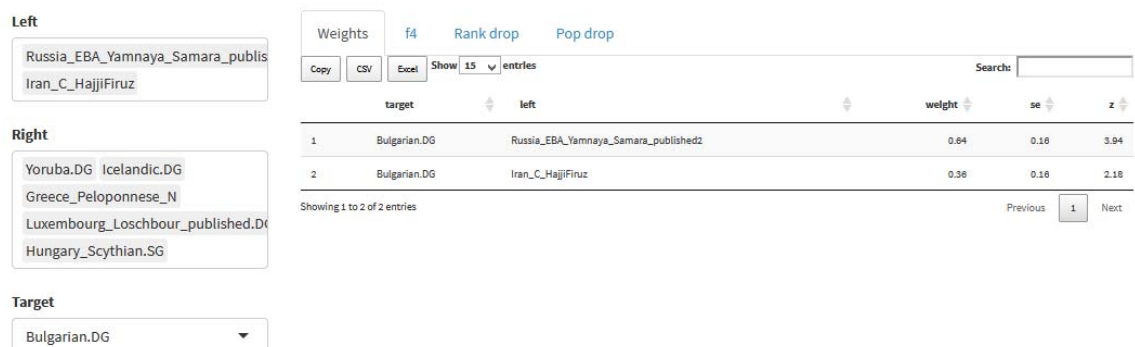
ADMIXTOOLS 2

| target | left | weight | se | z |
|----------------|--------------------------------------|---------------|-----------|----------|
| 1 Bulgarian.DG | Russia_EBA_Yamnaya_Samara_published2 | 0,85 | 0,14 | 6,12 |
| 2 Bulgarian.DG | Germany_EN_LBK_published | 0,15 | 0,14 | 1,07 |

In competition experiment with samples from European Neolithic, Anatolian Neolithic and Iranian Chalcolithic periods, performed with *qpAdm* we observed ambiguous results as samples from Western Iran Hajji_Firuz_Ch) on several occasions (depending on outgroup populations) performed better than European Neolithic samples in some of the competing models tested to identify the deep ancestry of contemporary Bulgarians; While contemporary Bulgarians exhibited

massive EEN component in all competing models, they could not be modeled as an admixture between Yamnaya and LBK, as both proportions and Z-scores from f4 statistics were not convincing in the presence of Iran_Hajji_Firuz_Ch sample, suggesting that Neolithic component of Bulgarians could not be modeled entirely with LBK-derived Neolithic component ($Z=1.07$), when combined with Yamnaya. When we tested Bulgarian population against different potential sources for their Neolithic component, we got the best fitting result when we substituted European LBK with Chalcolithic_Iran_HajjFruz sample and placed European LBK samples as an outgroup (Fig 4), (36% ancestry, Z-score $|2.18|$). We however caution that we set rather low threshold of 18 000 SNP due to the low resolution sequencing of some of the samples which undoubtedly had its effect and proportion results should not be taken at their face value .

Fig. 4



The results however suggested previously unreported deep chalcolithic Iran-related ancestry in modern Bulgarian population of unknown origin and unknown (due to the low-resolution samples) proportions.

The Chalcolithic sample from Iran is itself a mixture between Iranian and Anatolian Neolithic components. Previous studies have shown that ancient genomes from Neolithic and Copper Age farmers from the Iranian plateau harbored Anatolian farmer-related ancestry mixed with component from earlier herders of the western Zagros (The formation of human populations in South and Central Asia, *Narasimhan et al., 2019*). The particular sample from Hajji Firuz harbored 59% Anatolian farmer's related ancestry mixed with 41% Iranian Neolithic herders component. The languages of pre-copper age Iranian and Central Asian groups in unknown and we caution that in this survey we do not put linguistic meaning in the labels "Iran" and "Iranian".

Zagros-related Iranian farmers and herders reached Central Asia before 7000 YBC, and were widespread by 4000 YBC, when some of them adopted sedentary life style and began to settle in urban centers collectively known as BMAC. Iranian farmers/pastoralist had been mixing in Central Asia with the indigenous West Siberian hunter gatherers from an early point of their migrations as the inhabitants of BMAC already carried about 10% Western Siberian admixture. Pontic Steppe herders, related to the *Yamnaya* Steppe pastoralists arrived in Central Asia after 3000 YBC and added to this admixture

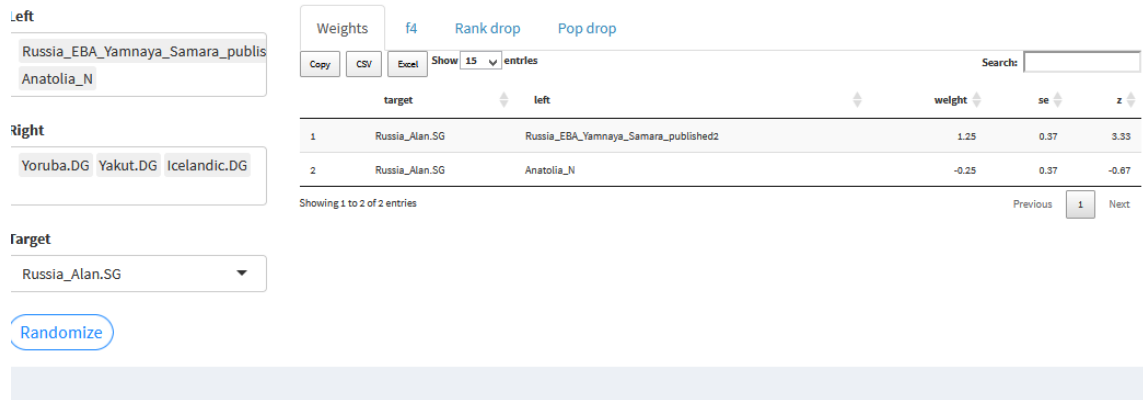
of farmers and hunter-gatherer a distinct Pontic Steppe component, having interacted with local populations on all levels. This interaction created number of tribes which became known to their neighbors under umbrella terms Saca, Scythians and Tocharians. Besides Yamnaya genetic component, they all had incorporated distinct ancestry from the Zagros Mountains and some ancestry from Siberian hunter gatherers, which made ancient Central Asian groups genetically distinct from ancient Europeans. Besides Yamnaya and Iran –related admixture, many Central Asian samples since Iron Age also harbored East Asian component.

Tracking the origins of the Central Asian signal in contemporary

Bulgarians:

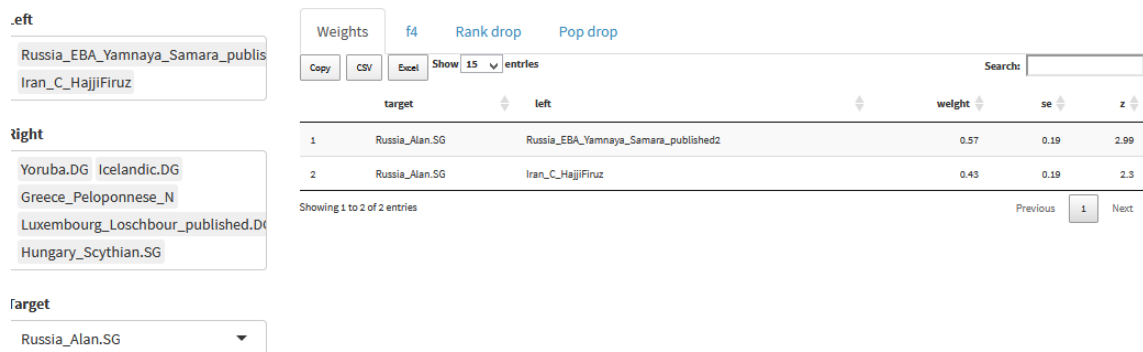
At first we employed *QPA* to test the samples from Saltovo Mayaki Culture as potential carriers of Central Asian component by competitively modeling them as an admixture of Yamnaya and Anatolian vs Zagros Neolithic farmers with the following results:

Fig. 5



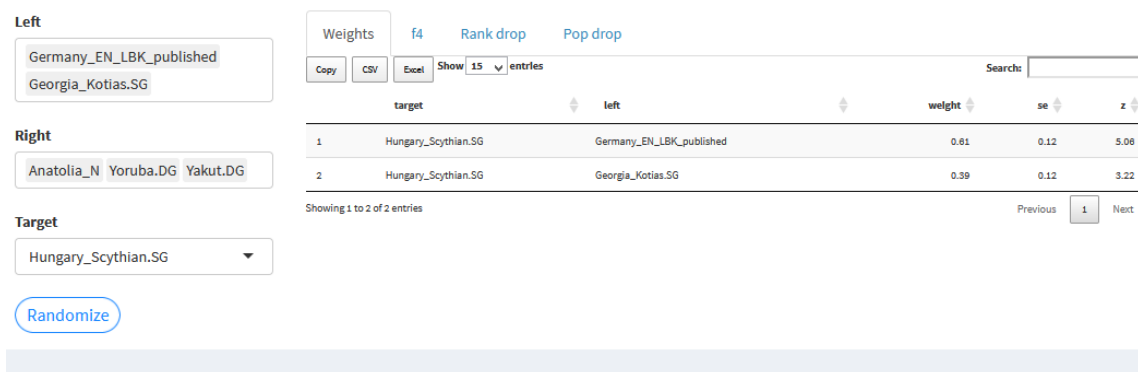
SMK population showed significant Yamnaya component ($Z=3.33$); However they did not exhibit clear European Neolithic ancestry, which suggested to a different place of origins than Europe. The above results have immediate implications on our understanding for the origins of Proto-Bulgarians, since majority of historians identify SMC with Old Great Bulgaria. While SMC population could not be modeled as a mixture of Yamnaya Neolithic (fig. 5), they however can be successfully modeled as 57% Yamnaya and 43% Iranian chalcolithic components. (Fig 6)

Fig. 6



We tested Late Antiquity Hungarian Scythian samples as a competing model for a potential source of the Neolithic Iran component in Bulgarians and in SMC. (Fig. 7). We established that Hungarian Scythian samples were lacking Iranian Neolithic heritage and hence could not be source of the Zagros component in the early medieval SMC and contemporary Bulgarians. To our surprise, the model with the best fit for the Hungarian Scythians included early Neolithic Caucasus sample (Kotias, Z |3.22|):

Fig. 7



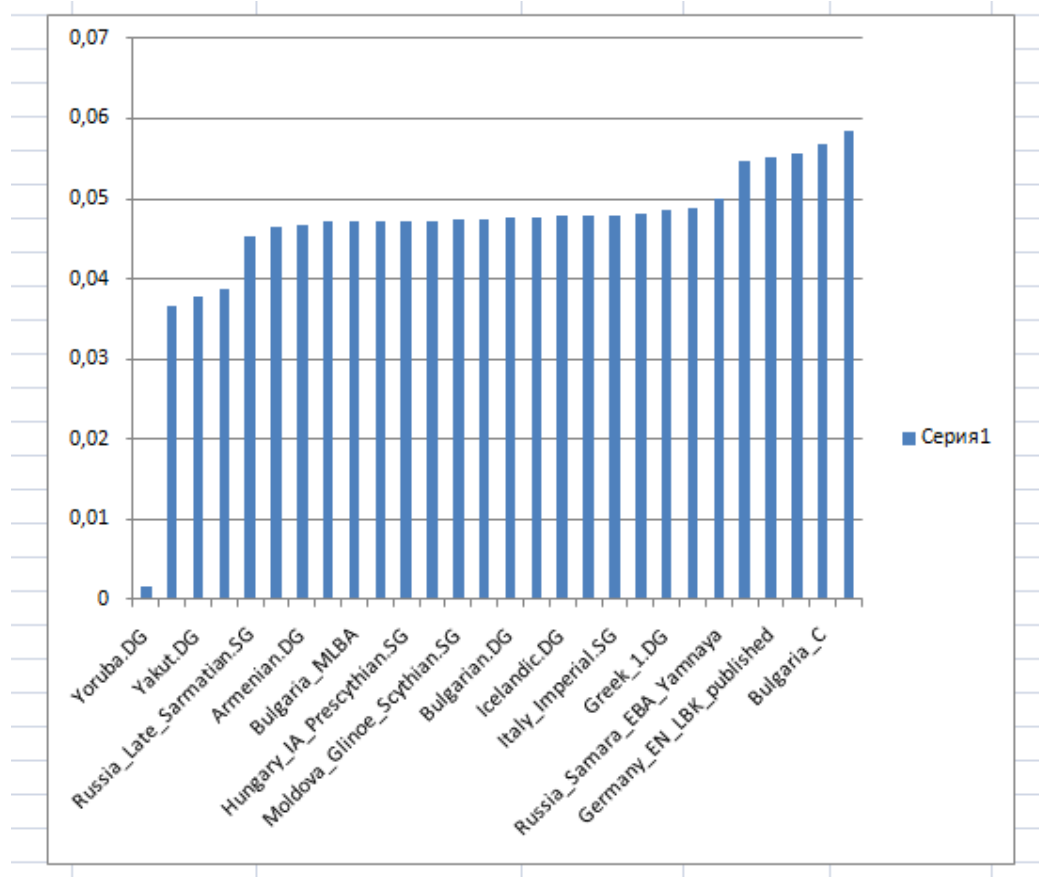
The results suggested Caucasian influence on the European Scythians from classical antiquity, but lack of connection between European Scythians and SMC ; the lack of Central Asian (chalcolithic Iran) component in European Scythians also suggested that they did not represent a back migration from Central Asia.

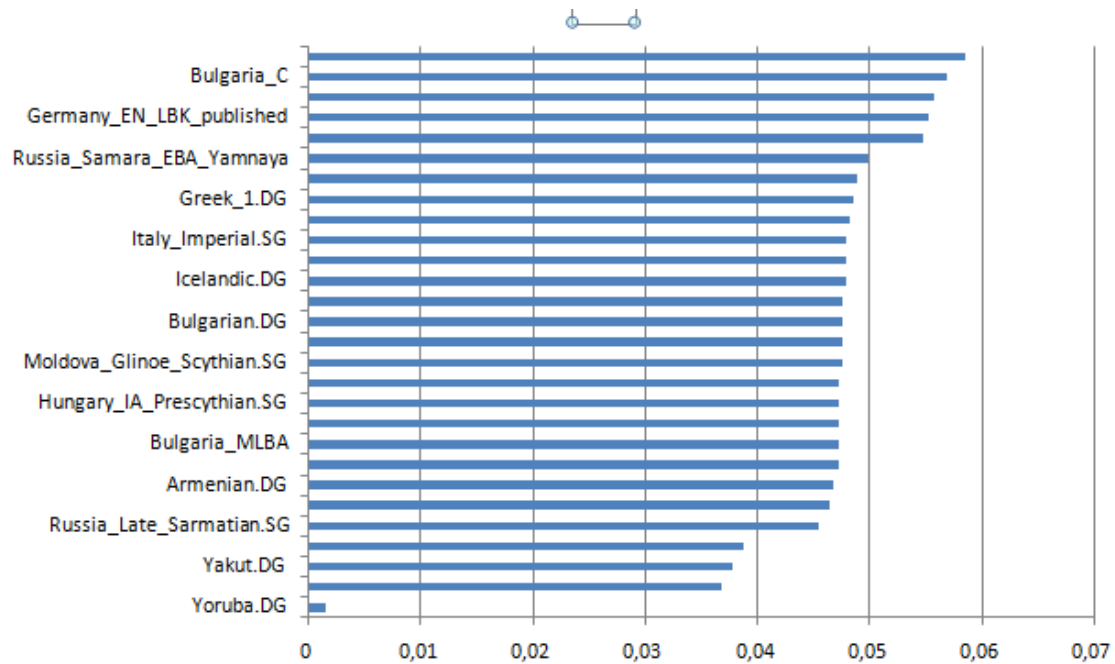
We next tested ancient Thracian samples for Iran-Neolithic ancestry and eventual relation to contemporary Bulgarians. After testing negative for Central Asian component, which excluded Thracians as a source, we computed f3 outgroup statistics for the Thracian samples from Mathieson et al (2018) in the form of (Dinka; IA_Bulgaria, X) against several populations from the past and several contemporary populations. We present the results in figures 8 and 9:

Fig. 8, f3 outgroup statistics:

| | | | | | | |
|----|----------|-------------|-------------------------|--------|----------|------|
| 27 | Dinka.DG | Bulgaria_IA | Yakut.DG | 0,0377 | 0,000908 | 41,5 |
| 19 | Dinka.DG | Bulgaria_IA | Karitiana.DG | 0,0388 | 0,000899 | 43,1 |
| 23 | Dinka.DG | Bulgaria_IA | Russia_Late_Sarmatian.S | 0,0454 | 0,000908 | 50 |
| 22 | Dinka.DG | Bulgaria_IA | Russia_Alan.SG | 0,0464 | 0,00089 | 52,1 |
| 2 | Dinka.DG | Bulgaria_IA | Armenian.DG | 0,0467 | 0,000917 | 51 |
| 1 | Dinka.DG | Bulgaria_IA | Armenia_LBA_LchashenN | 0,0472 | 0,000958 | 49,3 |
| 5 | Dinka.DG | Bulgaria_IA | Bulgaria_MLBA | 0,0473 | 0,00101 | 46,6 |
| 13 | Dinka.DG | Bulgaria_IA | Hungary_Avar_1.SG | 0,0473 | 0,00103 | 45,8 |
| 15 | Dinka.DG | Bulgaria_IA | Hungary_IA_Prescythian. | 0,0473 | 0,00102 | 46,2 |
| 26 | Dinka.DG | Bulgaria_IA | Ukraine_IA_WesternScyt | 0,0473 | 0,000951 | 49,7 |
| 20 | Dinka.DG | Bulgaria_IA | Moldova_Glinoe_Scythia | 0,0475 | 0,000891 | 53,3 |
| 21 | Dinka.DG | Bulgaria_IA | Polish.DG | 0,0475 | 0,000966 | 49,1 |
| 7 | Dinka.DG | Bulgaria_IA | Bulgarian.DG | 0,0476 | 0,000911 | 52,3 |
| 14 | Dinka.DG | Bulgaria_IA | Hungary_Avar_5 | 0,0476 | 0,00101 | 47,3 |
| 17 | Dinka.DG | Bulgaria_IA | Icelandic.DG | 0,0478 | 0,000947 | 50,5 |
| 8 | Dinka.DG | Bulgaria_IA | Cretan.DG | 0,0479 | 0,000903 | 53,1 |
| 18 | Dinka.DG | Bulgaria_IA | Italy_Imperial.SG | 0,0479 | 0,000844 | 56,8 |
| 16 | Dinka.DG | Bulgaria_IA | Hungary_IA_Scythian.SG | 0,0482 | 0,000917 | 52,6 |
| 11 | Dinka.DG | Bulgaria_IA | Greek_1.DG | 0,0485 | 0,000988 | 49,1 |
| 25 | Dinka.DG | Bulgaria_IA | Sardinian.DG | 0,0488 | 0,000889 | 54,9 |
| 24 | Dinka.DG | Bulgaria_IA | Russia_Samara_EBA_Yam | 0,0499 | 0,000887 | 56,2 |
| 4 | Dinka.DG | Bulgaria_IA | Bulgaria_EBA | 0,0548 | 0,00097 | 56,5 |
| 9 | Dinka.DG | Bulgaria_IA | Germany_EN_LBK_public | 0,0552 | 0,000904 | 61,1 |
| 10 | Dinka.DG | Bulgaria_IA | Greece_Peloponnese_N | 0,0557 | 0,000904 | 61,6 |
| 3 | Dinka.DG | Bulgaria_IA | Bulgaria_C | 0,0568 | 0,000884 | 64,2 |
| 6 | Dinka.DG | Bulgaria_IA | Bulgaria_N_published | 0,0585 | 0,000964 | 60,7 |

Fig 9 &10, f3 outgroup statistics for Iron Age Thracians versus selected populations:





The results suggest that ancient Thracians do not share more genetic drift with contemporary Bulgarians than with contemporary Icelanders, which we used as a referent population. On this ground, the results exclude substantial contribution to contemporary Bulgarians, coming from ancient Thracians. The results however suggest that ancient Thracians might have contributed more to of contemporary Greeks and to Latinized Balkan population from the Imperial Roman period. We noticed that IA Thracian samples shared most genetic drift with Neolithic and especially chalcolithic Balkan samples; they however did not show continuity with the preceding Copper and Bronze age samples from Bulgaria, which is puzzling.

Having failed to identify conclusive source of Central Asian admixture in modern Bulgarians in the nearby populations from the past, we computed 3-populations formal test of admixture for contemporary Bulgarians against *all* 2400 ancient samples from about 300 ancient Eurasian populations and ethnic groups we had available. In the figures below we report only the top results from 3-populations formal test of admixture for contemporary Bulgarians (negative Z-scores only). Surprisingly, the 3-populations test for admixture revealed an unexpected Wusun component in contemporary Bulgarians, as the top result from all pairs came when we combined Wusun samples with Early European Neolithic farmers.

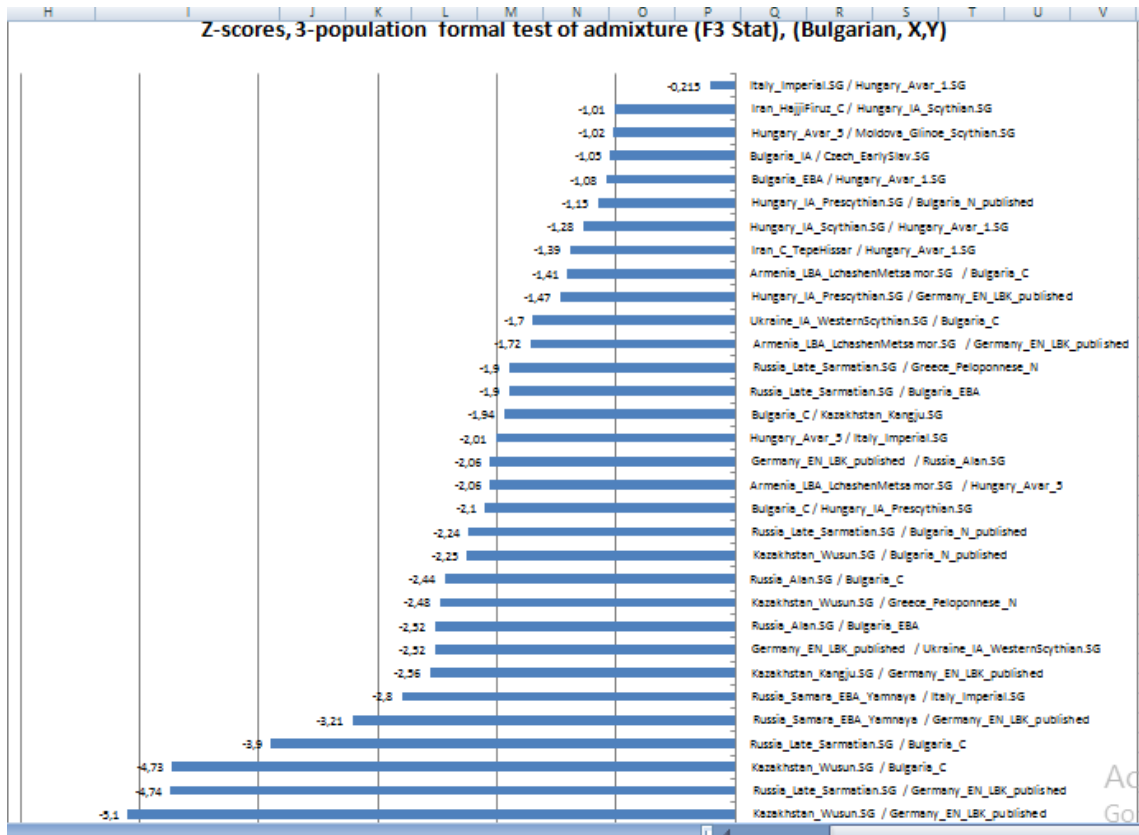
Fig. 11

| pop1 | pop2 | pop3 | est | se | 3-population formal test of p |
|--------------|--------------------------------|------------------------------|-----------|----------|-------------------------------|
| Bulgarian.DG | Kazakhstan_Wusun.SG | Germany_EN_LBK_published | -0.00235 | 0.000460 | -5,1 3.35e- 7 |
| Bulgarian.DG | Russia_Late_Sarmatian.SG | Germany_EN_LBK_published | -0.00204 | 0.000429 | -4,74 2.12e- 6 |
| Bulgarian.DG | Kazakhstan_Wusun.SG | Bulgaria_C | -0.00210 | 0.000443 | -4,73 2.25e- 6 |
| Bulgarian.DG | Russia_Late_Sarmatian.SG | Bulgaria_C | -0.00164 | 0.000421 | -3,9 9.71e- 5 |
| Bulgarian.DG | Russia_Samara_EBA_Yamnaya | Germany_EN_LBK_published | -0.00130 | 0.000405 | -3,21 1.34e- 3 |
| Bulgarian.DG | Russia_Samara_EBA_Yamnaya | Italy_Imperial.SG | -0.00105 | 0.000374 | -2,8 5.15e- 3 |
| Bulgarian.DG | Kazakhstan_Kangju.SG | Germany_EN_LBK_published | -0.00132 | 0.000515 | -2,56 1.06e- 2 |
| Bulgarian.DG | Germany_EN_LBK_published | Ukraine_IA_WestemScythian.SG | -0.00117 | 0.000463 | -2,52 1.16e- 2 |
| Bulgarian.DG | Russia_Alan.SG | Bulgaria_EBA | -0.00137 | 0.000544 | -2,52 1.18e- 2 |
| Bulgarian.DG | Kazakhstan_Wusun.SG | Greece_Peloponnese_N | -0.00117 | 0.000471 | -2,48 1.31e- 2 |
| Bulgarian.DG | Russia_Alan.SG | Bulgaria_C | -0.00109 | 0.000445 | -2,44 1.45e- 2 |
| Bulgarian.DG | Kazakhstan_Wusun.SG | Bulgaria_N_published | -0.00124 | 0.000553 | -2,25 2.43e- 2 |
| Bulgarian.DG | Russia_Late_Sarmatian.SG | Bulgaria_N_published | -0.00117 | 0.000521 | -2,24 2.52e- 2 |
| Bulgarian.DG | Bulgaria_C | Hungary_IA_Prescythian.SG | -0.00114 | 0.000544 | -2,1 3.59e- 2 |
| Bulgarian.DG | Armenia_LBA_LdashenMetsamor.SG | Hungary_Avar_5 | -0.00146 | 0.000710 | -2,05 3.98e- 2 |
| Bulgarian.DG | Germany_EN_LBK_published | Russia_Alan.SG | -0.000932 | 0.000452 | -2,05 3.89e- 2 |
| Bulgarian.DG | Hungary_Avar_5 | Italy_Imperial.SG | -0.000965 | 0.000480 | -2,01 4.44e- 2 |
| Bulgarian.DG | Bulgaria_C | Kazakhstan_Kangju.SG | -0.00100 | 0.000518 | -1,94 5.23e- 2 |
| Bulgarian.DG | Russia_Late_Sarmatian.SG | Bulgaria_EBA | -0.00108 | 0.000540 | -1,9 5.72e- 2 |
| Bulgarian.DG | Russia_Late_Sarmatian.SG | Greece_Peloponnese_N | -0.000841 | 0.000443 | -1,9 5.74e- 2 |
| Bulgarian.DG | Armenia_LBA_LdashenMetsamor.SG | Germany_EN_LBK_published | -0.000810 | 0.000472 | -1,72 8.61e- 2 |
| Bulgarian.DG | Ukraine_IA_WestemScythian.SG | Bulgaria_C | -0.000744 | 0.000438 | -1,7 8.95e- 2 |
| Bulgarian.DG | Hungary_IA_Prescythian.SG | Germany_EN_LBK_published | -0.000833 | 0.000567 | -1,47 1.42e- 1 |
| Bulgarian.DG | Armenia_LBA_LdashenMetsamor.SG | Bulgaria_C | -0.000671 | 0.000475 | -1,41 1.57e- 1 |
| Bulgarian.DG | Iran_C_TepeHassar | Hungary_Avar_1.SG | -0.000722 | 0.000519 | -1,39 1.64e- 1 |
| Bulgarian.DG | Hungary_IA_Scythian.SG | Hungary_Avar_1.SG | -0.000701 | 0.000549 | -1,28 2.02e- 1 |
| Bulgarian.DG | Hungary_IA_Prescythian.SG | Bulgaria_N_published | -0.000754 | 0.000556 | -1,15 2.51e- 1 |
| Bulgarian.DG | Bulgaria_EBA | Hungary_Avar_1.SG | -0.000686 | 0.000538 | -1,08 2.82e- 1 |
| Bulgarian.DG | Bulgaria_IA | Czech_EarlySlav.SG | -0.00180 | 0.00171 | -1,05 2.93e- 1 |
| Bulgarian.DG | Hungary_Avar_5 | Moldova_Glinoe_Scythian.SG | -0.000588 | 0.000577 | -1,02 3.09e- 1 |
| Bulgarian.DG | Iran_HajjiFiruz_C | Hungary_IA_Scythian.SG | -0.000541 | 0.000534 | -1,01 3.11e- 1 |
| Bulgarian.DG | Italy_Imperial.SG | Hungary_Avar_1.SG | -0.000101 | 0.000471 | -0,215 8.30e- 1 |

The results are puzzling and require further investigation. They however strongly suggest Wusun admixture in contemporary Bulgarians and point to Wusun tribes as carriers of the Iran Chalcolithic component, which we identified in contemporary Bulgarians. Considering the gradual increase in the European Neolithic component on the Balkans since Bronze age that we observed in our tests and the massive presence of Neolithic component in the Balkan samples from the Roman Imperial Ages (bigger than Neolithic component in IA Thracians), we propose that the Neolithic component detected in the f3 statistics

represents the genomic state of local Balkan population during the late antiquity. Besides Wusun and Mediterranean Neolithic component, contemporary Bulgarians expressed possible admixture signal from late Sarmatian samples from Pontic Caspian steppe and from Early Medieval Alans from SMC. We however identified Wusun as the carriers of the Central Asian signal in contemporary Bulgarians and we suggest possible relation between the migration of Proto-Bulgarians or Hunnic migration from Central Asia and the Wusun related component in contemporary Bulgarians. We detected same Wusun related component to a lesser extend in multiple populations from Eastern and Southeastern Europe (not reported in this survey), which raises the question if Wusun tribes formed the backbone of Hunnic migration in Europe.

Fig. 12



Very surprisingly, in our test Wusun samples outcompeted early Slav samples from Central Europe we had available. While results confirm ancestral contribution from Early Slavs, the evidence for Wusun admixture seem more convincing (Z-score $|-5.1|$ versus Z-scores $|-1.03|$).

Results from f-4 statistics re-confirmed the presence of Wusun – related component in contemporary Bulgarians:

Fig. 13

| | | | | | | z score |
|------------|-----------|--------------------------|---------------|------------|----------|---------|
| Bulgarian. | Yoruba.DC | Russia_Alan.SG | Dinka.DG | 0,0404 | 0,000482 | 83,8 |
| Bulgarian. | Yoruba.DC | Russia_Late_Sarmatian.SC | Dinka.DG | 0,04 | 0,000484 | 82,7 |
| Cretan.DC | Yoruba.DC | Russia_Alan.SG | Dinka.DG | 0,0402 | 0,000488 | 82,3 |
| Cretan.DC | Yoruba.DC | Russia_Late_Sarmatian.SC | Dinka.DG | 0,0392 | 0,000497 | 78,9 |
| Bulgarian. | Yoruba.DC | Kazakhstan_Wusun.SG | Dinka.DG | 0,0391 | 0,000521 | 75 |
| Bulgarian. | Yoruba.DC | Bulgaria_IA | Dinka.DG | 0,0415 | 0,000569 | 72,9 |
| Cretan.DC | Yoruba.DC | Bulgaria_IA | Dinka.DG | 0,0418 | 0,000574 | 72,8 |
| Cretan.DC | Yoruba.DC | Kazakhstan_Wusun.SG | Dinka.DG | 0,0382 | 0,000526 | 72,5 |
| Cretan.DC | Yoruba.DC | Russia_Alan.SG | Han.DG | 0,00871 | 0,000333 | 26,2 |
| Bulgarian. | Yoruba.DC | Russia_Alan.SG | Han.DG | 0,00857 | 0,000335 | 25,6 |
| Bulgarian. | Yoruba.DC | Russia_Late_Sarmatian.SC | Han.DG | 0,0082 | 0,000333 | 24,7 |
| Cretan.DC | Yoruba.DC | Russia_Late_Sarmatian.SC | Han.DG | 0,0078 | 0,000325 | 24 |
| Cretan.DC | Yoruba.DC | Bulgaria_IA | Han.DG | 0,0104 | 0,000452 | 23 |
| Cretan.DC | Yoruba.DC | Russia_Alan.SG | Yakut.DG | 0,00759 | 0,000335 | 22,6 |
| Bulgarian. | Yoruba.DC | Russia_Alan.SG | Yakut.DG | 0,00724 | 0,000337 | 21,5 |
| Bulgarian. | Yoruba.DC | Russia_Late_Sarmatian.SC | Yakut.DG | 0,00687 | 0,000333 | 20,6 |
| Bulgarian. | Yoruba.DC | Bulgaria_IA | Han.DG | 0,00961 | 0,000466 | 20,6 |
| Cretan.DC | Yoruba.DC | Russia_Late_Sarmatian.SC | Yakut.DG | 0,00667 | 0,000327 | 20,4 |
| Cretan.DC | Yoruba.DC | Bulgaria_IA | Yakut.DG | 0,00924 | 0,000454 | 20,3 |
| Bulgarian. | Yoruba.DC | Kazakhstan_Wusun.SG | Han.DG | 0,00724 | 0,000361 | 20 |
| Cretan.DC | Yoruba.DC | Kazakhstan_Wusun.SG | Han.DG | 0,00674 | 0,000361 | 18,7 |
| Bulgarian. | Yoruba.DC | Bulgaria_IA | Yakut.DG | 0,00828 | 0,000477 | 17,4 |
| Bulgarian. | Yoruba.DC | Kazakhstan_Wusun.SG | Yakut.DG | 0,0059 | 0,000381 | 15,5 |
| Cretan.DC | Yoruba.DC | Kazakhstan_Wusun.SG | Yakut.DG | 0,00562 | 0,000371 | 15,1 |
| Cretan.DC | Yoruba.DC | Bulgaria_IA | Bulgaria_MLBA | 0,00147 | 0,000449 | 3,27 |
| Cretan.DC | Yoruba.DC | Bulgaria_IA | Sardinian.DG | 0,000451 | 0,000376 | 1,2 |
| Bulgarian. | Yoruba.DC | Bulgaria_IA | Bulgaria_MLBA | 0,000504 | 0,000467 | 1,08 |
| Cretan.DC | Yoruba.DC | Russia_Alan.SG | Bulgaria_MLBA | -0,000191 | 0,000371 | -0,514 |
| Bulgarian. | Yoruba.DC | Bulgaria_IA | Sardinian.DG | -0,000373 | 0,000395 | -0,945 |
| Bulgarian. | Yoruba.DC | Russia_Alan.SG | Bulgaria_MLBA | -0,000541 | 0,000377 | -1,43 |
| Bulgarian. | Yoruba.DC | Russia_Late_Sarmatian.SC | Bulgaria_MLBA | -0,000911 | 0,000398 | -2,29 |
| Cretan.DC | Yoruba.DC | Russia_Late_Sarmatian.SC | Bulgaria_MLBA | -0,0011 | 0,00039 | -2,83 |
| Bulgarian. | Yoruba.DC | Kazakhstan_Wusun.SG | Bulgaria_MLBA | -0,00187 | 0,000441 | -4,25 |
| Cretan.DC | Yoruba.DC | Russia_Alan.SG | Sardinian.DG | -0,00121 | 0,000258 | -4,68 |
| Cretan.DC | Yoruba.DC | Kazakhstan_Wusun.SG | Bulgaria_MLBA | -0,00216 | 0,000427 | -5,06 |
| Bulgarian. | Yoruba.DC | Russia_Alan.SG | Sardinian.DG | -0,00142 | 0,000255 | -5,55 |
| Bulgarian. | Yoruba.DC | Russia_Late_Sarmatian.SC | Sardinian.DG | -0,00179 | 0,000291 | -6,15 |
| Cretan.DC | Yoruba.DC | Russia_Late_Sarmatian.SC | Sardinian.DG | -0,00212 | 0,000296 | -7,15 |
| Cretan.DC | Yoruba.DC | Kazakhstan_Wusun.SG | Sardinian.DG | -0,00318 | 0,00034 | -9,34 |
| Bulgarian. | Yoruba.DC | Kazakhstan_Wusun.SG | Sardinian.DG | -0,00275 0 | 0,000338 | -8.13 4 |

The results from f-4 statistics however suggested that this component could have arrived via SMC Alans or late Sarmatians from Russian plain. It is notable that despite geographical distance, Wusun had higher Z- scores than Iron Age Thracians.

Relationship between Bulgarians and ancient Mediterranean populations:

Fig. 14

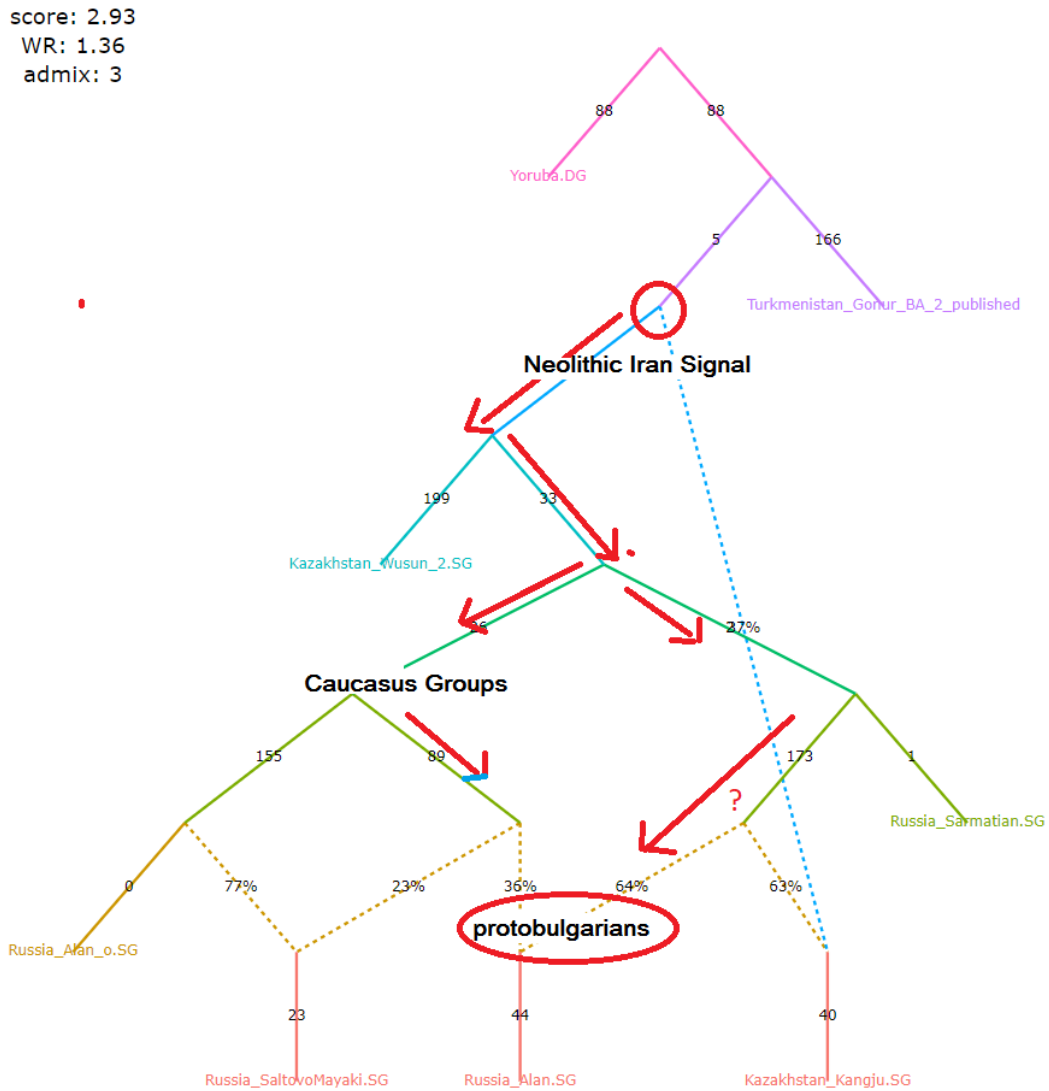
| | | | | | | | |
|----|-----------|-----------|-------------------------------|--------------|-----------|----------|--------|
| 5 | Bulgarian | Yoruba.DG | Italy_Imperial.SG | Dinka.DG | 0,0409 | 0,000441 | 92,7 |
| 9 | Bulgarian | Yoruba.DG | Bulgaria_C | Dinka.DG | 0,042 | 0,000454 | 92,7 |
| 21 | Bulgarian | Yoruba.DG | Moldova_Glinoe_Scythian.SG | Dinka.DG | 0,0413 | 0,000472 | 87,5 |
| 17 | Bulgarian | Yoruba.DG | Hungary_IA_Scythian.SG | Dinka.DG | 0,0415 | 0,000489 | 84,9 |
| 25 | Bulgarian | Yoruba.DG | Ukraine_IA_WesternScythian.SG | Dinka.DG | 0,0407 | 0,000486 | 83,8 |
| 13 | Bulgarian | Yoruba.DG | Bulgaria_N_published | Dinka.DG | 0,0412 | 0,000498 | 82,6 |
| 1 | Bulgarian | Yoruba.DG | Bulgaria_IA | Dinka.DG | 0,0415 | 0,000569 | 72,9 |
| 7 | Bulgarian | Yoruba.DG | Italy_Imperial.SG | Han.DG | 0,00906 | 0,000293 | 30,9 |
| 11 | Bulgarian | Yoruba.DG | Bulgaria_C | Han.DG | 0,0102 | 0,00033 | 30,9 |
| 23 | Bulgarian | Yoruba.DG | Moldova_Glinoe_Scythian.SG | Han.DG | 0,00943 | 0,000319 | 29,5 |
| 19 | Bulgarian | Yoruba.DG | Hungary_IA_Scythian.SG | Han.DG | 0,0097 | 0,000357 | 27,1 |
| 27 | Bulgarian | Yoruba.DG | Ukraine_IA_WesternScythian.SG | Han.DG | 0,00889 | 0,000339 | 26,2 |
| 10 | Bulgarian | Yoruba.DG | Bulgaria_C | Yakut.DG | 0,00886 | 0,000341 | 26 |
| 22 | Bulgarian | Yoruba.DG | Moldova_Glinoe_Scythian.SG | Yakut.DG | 0,0081 | 0,000323 | 25,1 |
| 6 | Bulgarian | Yoruba.DG | Italy_Imperial.SG | Yakut.DG | 0,00773 | 0,000313 | 24,7 |
| 15 | Bulgarian | Yoruba.DG | Bulgaria_N_published | Han.DG | 0,00935 | 0,000382 | 24,5 |
| 18 | Bulgarian | Yoruba.DG | Hungary_IA_Scythian.SG | Yakut.DG | 0,00836 | 0,000352 | 23,8 |
| 26 | Bulgarian | Yoruba.DG | Ukraine_IA_WesternScythian.SG | Yakut.DG | 0,00756 | 0,000345 | 21,9 |
| 3 | Bulgarian | Yoruba.DG | Bulgaria_IA | Han.DG | 0,00961 | 0,000466 | 20,6 |
| 14 | Bulgarian | Yoruba.DG | Bulgaria_N_published | Yakut.DG | 0,00802 | 0,00039 | 20,6 |
| 2 | Bulgarian | Yoruba.DG | Bulgaria_IA | Yakut.DG | 0,00828 | 0,000477 | 17,4 |
| 12 | Bulgarian | Yoruba.DG | Bulgaria_C | Sardinian.DG | 0,000211 | 0,000221 | 0,955 |
| 4 | Bulgarian | Yoruba.DG | Bulgaria_IA | Sardinian.DG | -0,000373 | 0,000395 | -0,945 |
| 20 | Bulgarian | Yoruba.DG | Hungary_IA_Scythian.SG | Sardinian.DG | -0,000289 | 0,000264 | -1,1 |
| 16 | Bulgarian | Yoruba.DG | Bulgaria_N_published | Sardinian.DG | -0,000636 | 0,00029 | -2,19 |
| 24 | Bulgarian | Yoruba.DG | Moldova_Glinoe_Scythian.SG | Sardinian.DG | -0,000551 | 0,000247 | -2,23 |
| 8 | Bulgarian | Yoruba.DG | Italy_Imperial.SG | Sardinian.DG | -0,000928 | 0,000199 | -4,65 |

We conducted multiple tests and got highest Z score from Roman Imperial samples, Moldova Scythians and Bulgarian Chalcolithic samples, all of which showed convincing presence in Bulgarian genomes; we also detected hints for East Asian component.

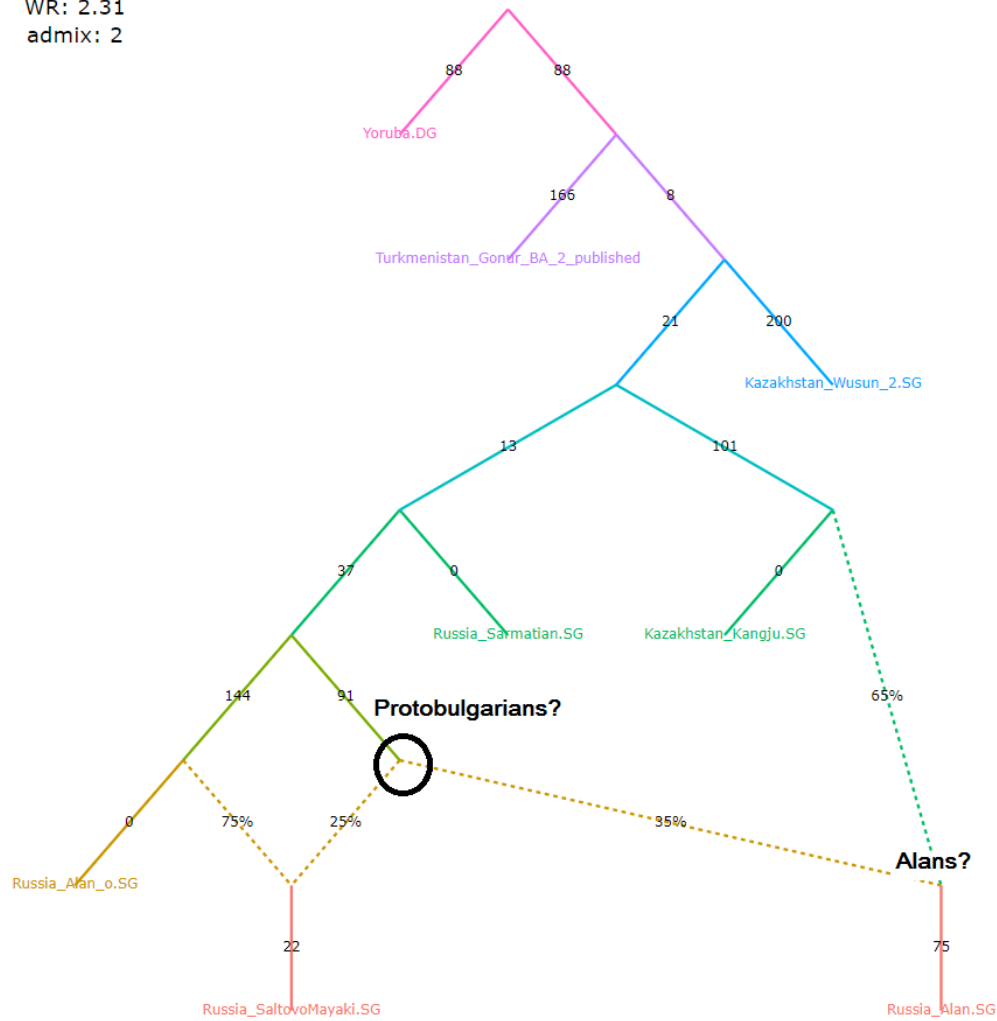
Modeling the arrival of Iranian Neolithic ancestry in Early Medieval Pontic Steppe and Early Medieval Balkans.

We used *qpGraph* to model the arrival of Iranian Neolithic ancestry in Medieval Pontic Steppe and as we arrived to two equally plausible models. We selected the samples in the graphics based on their F-3 and F-4 statistics results.

Fig. 15



score: 12.01
WR: 2.31
admix: 2



Discussion

We caution that in this survey we tracked only the origins of Central Asian ancestral component in modern Bulgarians. While we did not evaluate the exact proportions of East Asian and Siberian ancestry in Wusun, SMC and contemporary Bulgarian samples, we detected that most surveyed samples carried varying degree of it as according to the results from f4 statistics, both

Wusun samples and SMC samples showed presence of East Asian (SMK) and Western Siberian (Wusun) ancestral component. To our knowledge the only sequenced sample from First Bulgarian Kingdom also expressed certain amount of East Asian ancestry (D. Reich, unpublished, personal communication). With DNA sequencing of more Proto-Bulgarian samples from First Bulgarian Kingdom already underway, we hope that we will soon be able to give an estimate of East Asian ancestry in both Proto-Bulgarians and in contemporary Bulgarians.

While we included 3 Early Medieval Slavic samples in our survey they proved insufficient to clarify the relationship between Early Slavs, Early Medieval Steppe Nomads and contemporary Bulgarians. Guided by the results from f-3 and f-4 statistics, we suggest a substantial interaction between Slavonic tribes and migrating steppe nomads from Central Asia. One possible interpretation of the Wusun component in contemporary Bulgarians would be the early presence of Wusun Component in the migrating Early Slavs that might have emerged during their well documented military alliances with European Huns, Pannonian Avars and Proto-Bulgarians. If this would be the case, the carriers of the Wusun ancestry in modern Bulgarians would be early Slavic tribes that had migrated to Balkan Peninsula during Early Medieval Ages.

References

- Rashev, R. Prabylgarite prez 5-ti – 7-mi vek, (2005)
- Haak W et al., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. (2015).
- 8Allentoft ME et al., Population genomics of Bronze Age Eurasia. *Nature* (2015).
- Lazaridis I et al., Genomic insights into the origin of farming in the ancient Near East. *Nature* (2016).
- Mathieson I et al., The genomic history of southeastern Europe. *Nature*. (2018).
- Lipson M et al., Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*.
- Lazaridis I et al., Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. (2014).
- Gallego-Llorente M et al., The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep* 6, (2016).
- B. Damgaard P et al., 137 ancient human genomes from across the Eurasian steppes. *Nature*. (2018).
- Barros Damgaard P et al., The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*. (2018).
- Lazaridis I et al., Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* (2014). [
- Patterson N et al., Ancient admixture in human history. *Genetics*. (2012).

Fernandes, D.M., Mittnik, A., Olalde, I. *et al.*, The Arrival of Steppe and Iranian Related Ancestry in the Islands of the Western Mediterranean (2020)

The formation of human populations in South and Central Asia, *Narasimhan et al.*, (2019)