bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Genetic evidence suggests relationship between contemporary Bulgarian population and Iron Age steppe dwellers from Pontic-Caspian steppe.

Todor Chobanov Ph.D.,Ass.prof., Bulgarian Academy of Science, Svetoslav Stamov MA , Duke University

Abstract

Ancient DNA analysis on the ancestry of European populations conducted in the last decade came to the puzzling conclusion that while all contemporary European populations can be best represented as an admixture of 3 ancestral populations –Early European Neolithic farmers (ENF), Western Hunter-Gatherers (WHG) and Ancestral North Eurasians (ANE), contemporary Bulgarians and few other SEE populations can also be represented as an admixture of two groups only – Early European Neolithic farmers and contemporary Caucasian people equally well.

If modeled as an admixture of two groups only, the ANE component presented in contemporary Bulgarians would have arrived on the Balkans with hypothetical ANE (Ancestral North Eurasians)-rich Caucasian population.

In this paper, we test the hypothesis that increased Caucasian component in contemporary SE Europeans, has been introduced on the Balkans by migrating Iron Age steppe dwellers from Pontic-Caspian steppe. We analyze available DNA datasets from both ancient and contemporary samples and identify a Caucasian signal, carried to Balkan populations by the nomadic dwellers of IA Saltovo-Maiaki Culture, located on the northern slope of Caucasus Mountains and adjacent steppe regions. We also identify two additional sources of Caucasian admixture in SEE populations, which are not specific to Bulgarian population only. Based on the results from our population genetic analysis we

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

suggest that contemporary Bulgarians are an admixture of ancestral Slavonic groups, rich on locally absorbed EEF DNA and Proto Bulgarians, rich on Caucasian DNA and genetically related to the bearers of the Saltovo-Mayaki Culture from 6-8 century AD.

Introduction

All contemporary European populations can be represented as an admixture of 3 ancient groups: Early European Neolithic farmers (ENF), western hunter-gatherers (WHG) and Ancestral North Eurasians (ANE). (Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409-13.)

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

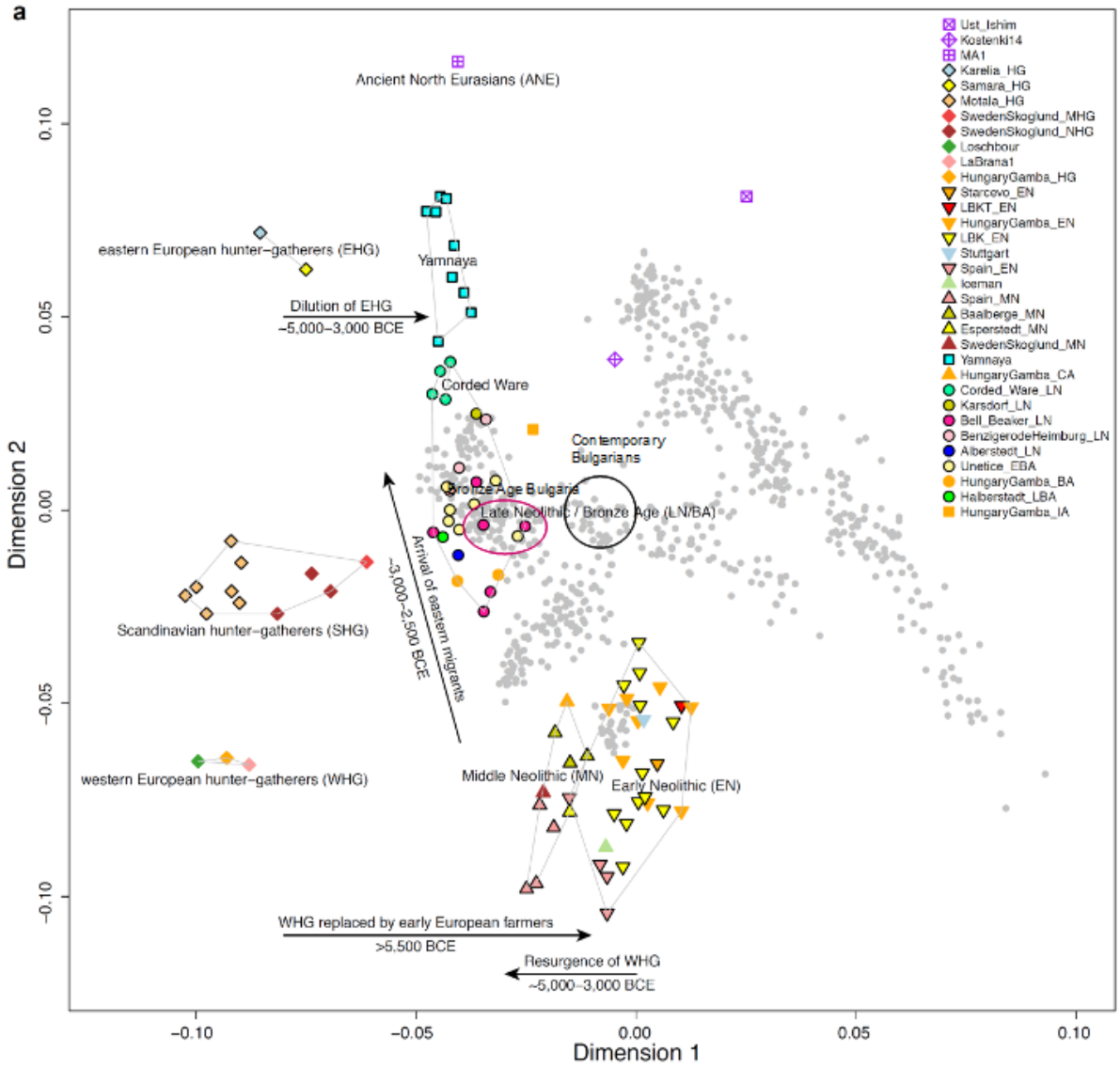

Fig. 1 Contemporary Bulgarians show an extra layer of Caucasian admixture, which is missing from the Bronze Age Balkan population (BAB). BAB are a mixture of Yamna migrants and EEF – just as rest of European populations. On the plot we can see that contemporary Bulgarians are closer to the Caucasian cluster than Bronze Age Balkan samples are. PCA after Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015;522(7555):207-11., Mathieson I, Alpaslan-Roodenberg S, Posth C, et al. The genomic history of southeastern Europe. Nature. 2018;555(7695):197-203

On the map (Fig. 1) contemporary Bulgarians are distributed nearer to contemporary Caucasians than most European populations which suggests an extra degree of Caucasian admixture that has been absent in the rest of Europe. This implies admixture events that are specific to Bulgarian population and whose effects are limited to the area of Balkan Peninsula mostly.

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Laazridis and Reich first noted that like the rest of Europeans, south-east Europeans can best be modeled as 3-way admixture (ANE-EEF-WHG), however they can be modeled as 2-way only admixture equally well (EEF-Caucasians, where ANE component would have come from additional Caucasian migrations to the Balkans). Haak et al confirmed the findings of D. Reich and established a vector of massive migration from Black Sea – Caspian steppe region into Europe. This migration occurred during early Bronze Age and became major contributing factor to the populations of all contemporary Europeans. Haak established that BA migrants represented an admixture of Caucasian Hunter Gatherers, genetically rooted in Mesolithic Northern Iran and East European Hunter Gatherers from what is now Russian plain. The migrants carried distinctive Caucasian signature and introduced Caucasian component throughout European continent. While this signature had been dilated in Western Europe in the centuries that followed, it had increased in the Balkan populations. This increase is suggestive of more admixture events with populations, caring Caucasian component and limited to the Balkans only. (see Fig. 1.)

The increase in Caucasian component in contemporary Bulgarians postdates Bronze Age migrations. Historical literature suggests that the arrival of this component in Bulgarian population could be related to the migration of Protobulgarians (Bulgars) during 6-8 century AD and the foundation of First Bulgarian Kingdom (V. Zlatarski, S. Runsiman, R.Rashev). Century long archeological research has identified northern Caucasian slopes and adjacent Kuban River zone as the likely homeland of the migrating Bulgars. Archaeological research suggests intensive contacts between Bulgars and the neighboring Caucasian and Alanic tribes, including the emergency of material culture of mixed origin, suggestive of a synthesis between IA Caucasian and IA

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

steppe traditions, emerging in the zone of Cuban river during early IA Saltovo-Maiaki Culture (SMC, 6-8 century AD). In this paper, we present the results from our analysis on the available ancient genetic data from BA and IA Western Eurasia, including samples from SMC in their relation to modern Bulgarians.

Method

We analyzed ancient DNA samples from Bronze Age, Iron Age and medieval Western and Central Eurasia. In an attempt to establish the source population and the timing of the additional Caucasian admixture in contemporary Bulgarians, we merged the ancient dataset with the dataset of 40 contemporary Bulgarians as well as the dataset of 100 contemporary individuals from neighboring populations. We computed principal component analysis on the present populations and projected available ancient DNA samples from Western and Central Eurasia. We also built a neighbor joining tree of the available ancient and contemporary samples. All genetic trees and PCA plots have been computed with PAST software for palaeogenetic DNA analysis.

We also reviewed already published genetic research on the topic in the scientific literature in order to identify what has been already known about the timing and the hypothesized source population. We also test several well-known historical hypothesizes about the origins of contemporary Bulgarians and early IA Protobulgarians.

Results

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Using statistical genome-wide analysis, we detected nontrivial genetic connection between contemporary Bulgarians, inhabitants of Bronze Age Armenian plateau and Iron Age dwellers from SMC. Our analysis also suggests surprising connection between contemporary Bulgarians and Iron Age Scythians from Hungarian plain.

Principal Component Analysis

For our PCA and genome-wide statistical analysis we used PAST3.22, version December 2018 - Paleontological statistics software package for education and data analysis (Hammer 2001).

All contemporary individual DNA genome-wide data files were retrieved from Yunusbayev et al 2012. To analyze the genetic distances and genetic relationship of the retrieved samples to the contemporary Bulgarian samples, we built several principle component analysis (PCA) plots, which visualized the genetic relationship between the individuals, their genetic contribution to the contemporary Bulgarians and we created several genetic trees based on their degree of relatedness.

In our first PCA (Fig 2) we combined dataset from 137 ancient samples from the Eurasian Steppe - from what is now Mongolia to what is now Hungarian plain (P. Damgaard et al, Nature volume 557, pp369–374, May 2018) and merged it with selected contemporary individuals from SE Europe (dataset from Yunusbayev et al 2012)

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

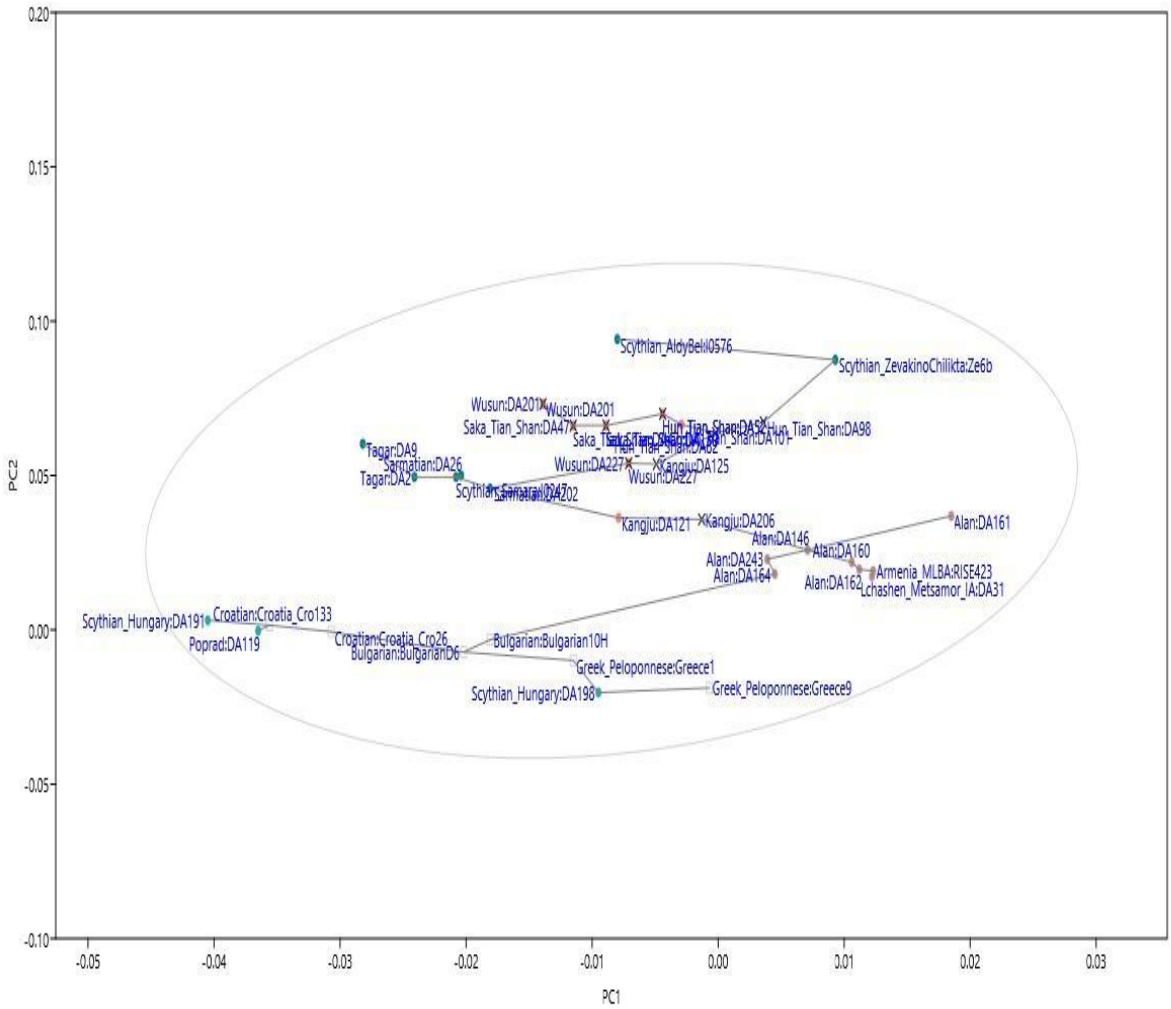

Fig. 2 PCA on the relationship between contemporary Bulgarians and ancient samples from BA and IA Eurasian steppe. While none of the contemporary Bulgarians yields relation to the ancient CA populations, PCA1 suggests genetic connection between contemporary Bulgarians and IA individuals ***AlanDA243, AlanDA164 and Alan DA146*** from North Ossetia and SMC.

The results of PCA (Fig 2) renders direct connection between contemporary Bulgarians and Inner Asian steppe nomads from migration period unlikely. None of the contemporary Bulgarians yielded any direct or mediated relation to the ancient Far Eastern and Central Asian nomadic steppe populations.

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In order to examine population transformation in what is now contemporary Bulgaria from early Bronze Age trough Iron Age till now, we also added 8 ancient samples from the late Neolithic / Early Bronze Age and early Iron Age, which we retrieved from Haak et al 2015, 207-11 and from Mathieson et al 2018, 197-203.). We present the results in Fig. 3

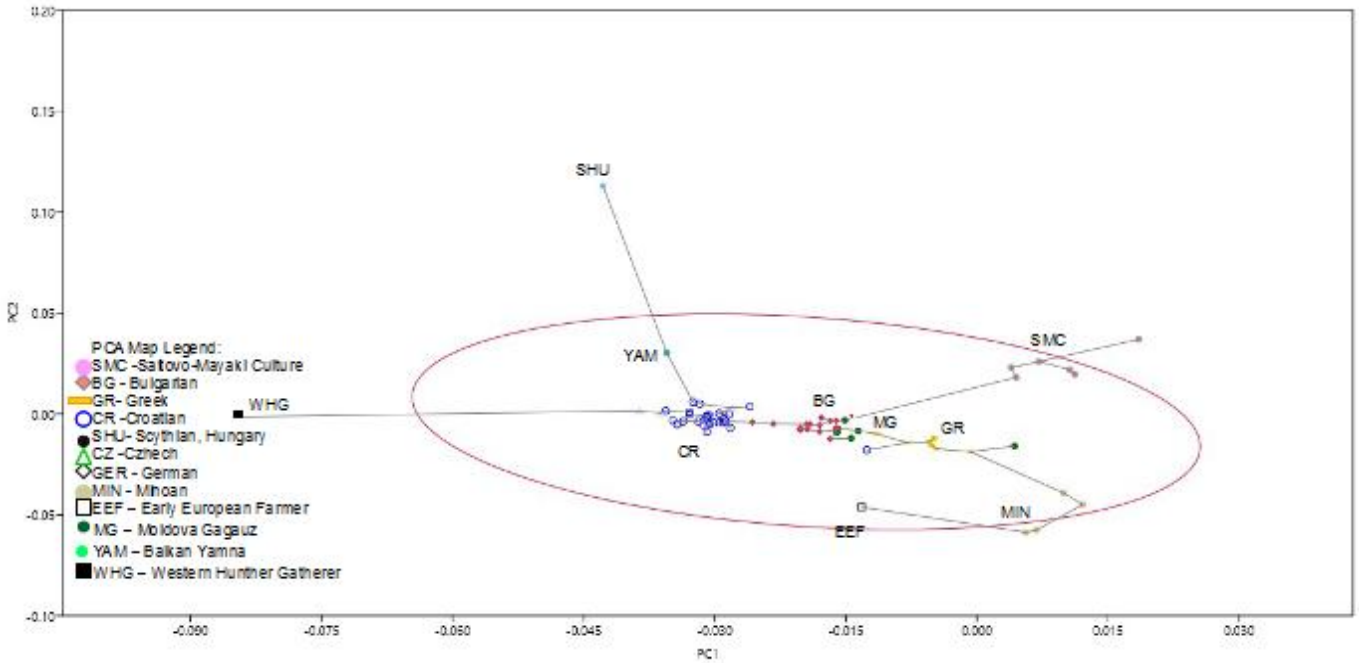

Fig 3 There is statistically significant relationship between contemporary Bulgarians and the Protobulgarians from SM. The genetic affinities detected by PAST3 suggest that SM people have contributed to contemporary Bulgarians only and their contribution to the rest of Balkan population has been transmitted from contemporary Bulgarians to their geographical neighbors.

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PCA results suggest genetic connection between contemporary Bulgarians and the ancient individuals AlanDA243, AlanDA164 and Alan DA146 belonging to SM culture.

In our next PCA we added Scythian samples from Hungarian plain from 4th Century BC (classical antiquity). The plot suggests connection between Scythian samples, European Alans from the migration period and the nomads from the *Saltovo-Mayaki* Culture as all 3 groups showed genetic connection to contemporary Bulgarians. (fig. 4)

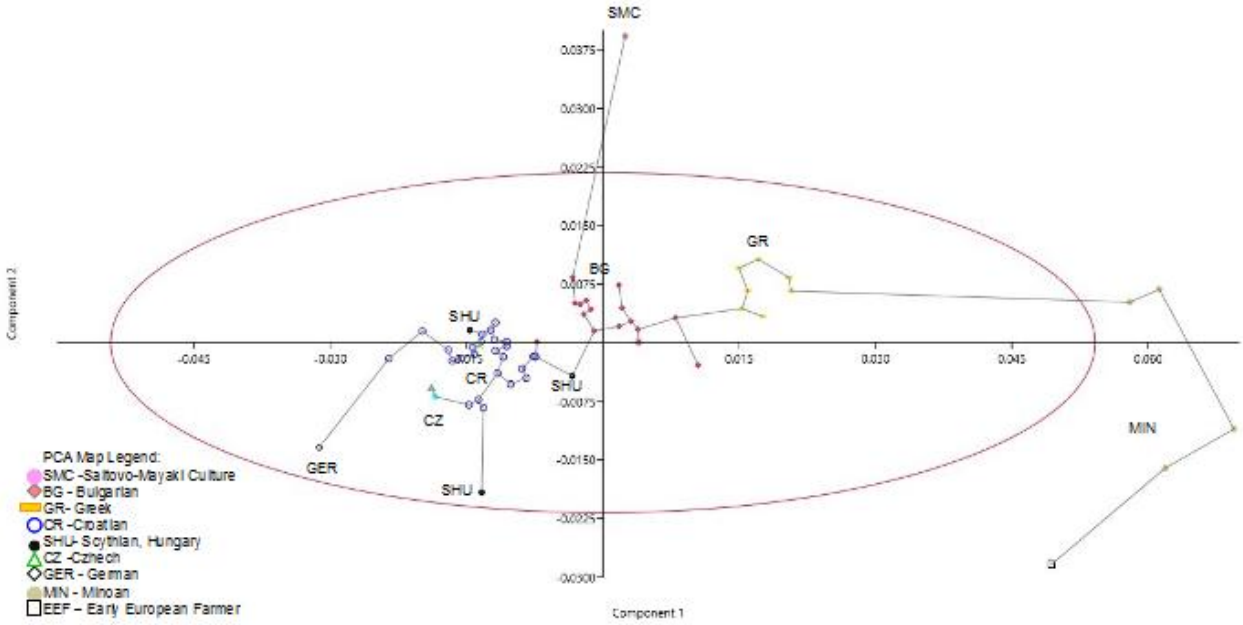

Fig. 4

These results imply nomadic influence from migration period being carried over to the population genomics of contemporary Bulgarians.

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Our PCA (Fig.2) also revealed indirect connection between contemporary Bulgarians and central Asian Bronze Age nomads of East Iranic origin known as Kangju group. This relation however is dependent on the presence of sample Alan DA146 from *Saltovo-Mayaki* (*Saltovo*, SM for short) culture on the PCA Plot and disappears if we remove this sample from the plot. We suggest that this discrete connection represents earlier stages of the migration of certain proto SM groups (Sarmatians-Alans?). Yet the rest of SM samples did not yield same connection to Kangju but showed detectable connection to the samples from Bronze Age Armenian plateau (fig. 2), suggestive of multiple admixture events during different earlier stages of migrations and contacts of SM people, as one of these stages must have included Armenian plateau in Central Caucasus.

Since there were multiple waves of migration from Caucasus to the Balkans including IE migration during Bronze Age and the emergence of Minoans during early BA and they all carried substantial Caucasian component with them (Haak et al 2015,207-11. and Mathieson et al 2018, 197-203), in our next plot we tried to distinguish the admixture signal coming from SM people from admixture signals coming from the earlier migrations. In order the test the Huns as potential carriers of the same signal, we also included a sample of iron-age Siberian hunter gatherer as a proxy for the Huns and in order to test the early Slavs for yet another potential carrier, we included contemporary Croatian samples as a proxy for the medieval Bulgarian Slavs. We also included Moldova Gagauz samples to test if they carry stronger Protobulgarian signal as it has been hypothesized by some of Bulgarian historians. We present the results in Fig. 5 and Fig. 6:

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

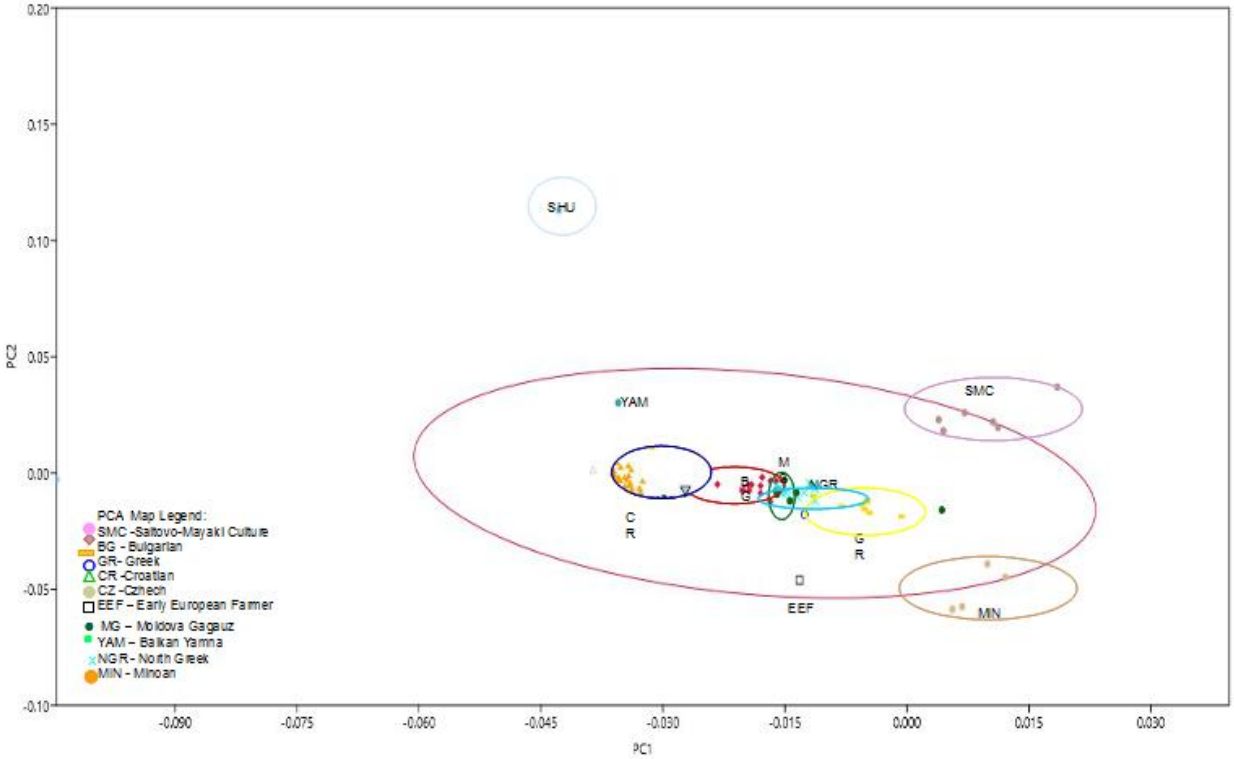

Fig. 5

In the PCA plot (fig. 5) the current Balkan nations form a cline. None of the tested samples showed detectable relation to SHG sample. The signals coming from SMC, NPBA Minoans and Bronze Age proto Thracians are clearly distinguishable from each other. Moldova Gagauz samples take intermediary position between contemporary Bulgarians and Contemporary Greeks and do not show stronger connection to SMC than contemporary Bulgarians, hence the signal from Protobulgarians in contemporary Bulgarians comes directly from SM and is not mediated by Gagauz people (which also carry this signal). Bronze Age proto Thracians are genetically closer to early medieval Slavs (represented here by Croatian samples) than to contemporary Bulgarians and

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

their influence on Bulgarian population genomics is not direct, but is probably mediated by early Slavs;

Peloponnese Greeks show closest affinity to Neolithic Peloponnesus and Bronze Age Minoans (fig. 5 and fig. 6). We conclude that the influence of Minoans on contemporary Bulgarian population is not direct and is due to population transfers and exchanges that led to admixture between medieval Bulgarians, medieval Greeks and medieval ERE populations. Both contemporary Greeks and contemporary Bulgarians show considerable distance to Bronze Age Balkan Yamna population (Thracians?) and Thracian contribution is mediated by the Croatians (fig. 5) as a proxy of the early Slavs, unless it masks Illyrian contribution in contemporary Croatians. We cannot determine whether Croatian samples reflect Illyrian or Thracian influence on the genomes of early Slavs based on the available data only. Further research is needed to clarify this topic.

We noted that SM (Protobulgarian-Alan) influence among contemporary Balkan nations has its strongest representation in contemporary Bulgarians (Fig 4) where it arrives directly and this Protobulgarian influence in the other Balkan nations is mediated by the contemporary Bulgarians who channel it.

Neighbor joining tree, built with PASTX software on the base of genetic relationship between the samples:

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

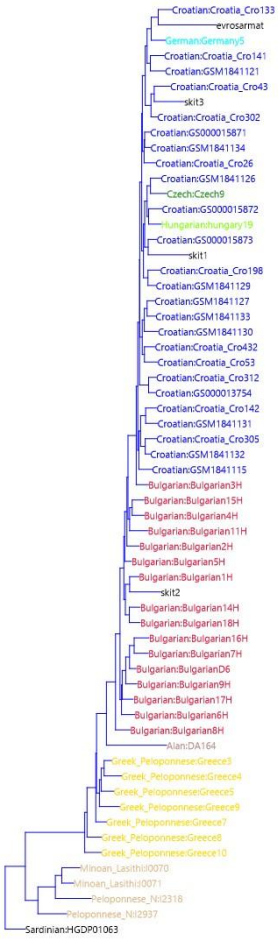

FIG 6 . Neighbor joining tree

Conclusions from the DNA data analysis

The results suggest that SMC related populations are among the precursor of contemporary Bulgarians. This makes SM culture at its precursor stage (600-700 AD) leading candidate for the source population of Asparukh Bulgarians. These results also suggest that Asparukh's tribe(s) are indistinguishable from the Sarmato-Alanic groups from Early MA and Late antiquity and, surprisingly, do not carry Siberian and Central Asian admixture on the Balkans with them.

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Unlike BA Thracians and the early Slavs, SMC carry substantial Caucasus admixture, related to the tribes from Bronze Age Armenian plateau and seems to have transmitted this admixture to the contemporary Bulgarians. The relationship between Protobulgarians and Sarmato-Alanic tribes from the Late antiquity and Early medieval epoch remains to be clarified further, however genome wide-data suggest that Protobulgarians were themselves an admixture in equal proportions between two close, but distinct populations –1. Alano-Sarmatian tribe from the region north of Caucasus with some Kangju link to it and 2. Unknown tribe(s) originating from what is now Armenian Plateau. Both Scythian samples from the Hungarian steppe and the Alans from *Saltovo-Mayaki* culture bear strong genetic resemblance to the Bronze Age Caucasian samples, which is missing in central Asian nomads but is presented in the contemporary Bulgarians.

Our results cast a doubt on a connection between Inner Asian nomadic tribes from Antiquity and the Protobulgarians-Alans from SM culture and Northern Caucasus. The lack of Inner Asia autosomal DNA links for the Protobulgarians confirms the results from the mtDNA sampling of materials from 8th-9th c. necropolises on the Lower Danube. The main haplogroup H (H, H1, H5, and H13) prevalent in European populations has a 41.9% frequency in modern Bulgarians, and it was observed in 7 of 13 proto-Bulgarian samples. Again no evidence was found of East Asian (F, B, P, A, S, O, Y, or M derivative) haplogroups (Nesheva et al 2015, 22). An earlier major representative survey of present dale male lineages in Bulgaria (over 800 individuals) revealed that "*Haplogroups C, N and Q, distinctive for Altaic and Central Asian populations, occur at the negligible frequency of only 1.5%.*" (Karachanak et al 2013). Our research suggest that author's conclusion of the survey that *"…our data suggest that a common paternal ancestry between the proto-Bulgarians and*

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the Altaic and Central Asian populations either did not exist or was negligible..."(Karachanak et al 2013, abstract) was correct.

Since the debate about potentially "autochthonous" component in the contemporary Bulgarians (present day version of "Illyrism") has become somewhat hotly debated topic in Bulgarian society today, we also clarified the origin of this Caucasian component further and managed to split the Caucasian component coming from SM from the Caucasian components already presented on the Balkans prior to Protobulgarian migration. We established that while all three carry somewhat similar Caucasian component (fig.3, fig.4, fig.5), the signal, coming from SM is the strongest in contemporary Bulgarians, the signal coming from Bronze Age Thracians is the strongest in contemporary Croatians and the signal, coming from Bronze Age Minoans is the strongest in contemporary Greeks. These three signals clearly differ from each other and their source populations are clearly distinguishable. Yet all tree carry an excessive Caucasian component, suggesting non-local origins for all three of them and suggestive of at least three different migrations from the Caucasus Mountains to the Balkans. However, contemporary Bulgarians have received their Minoan component mostly through population exchange with Byzantium and their Bronze age Thracian component trough admixture/population exchange with early medieval Slavs and Croats. The signal that distinguished contemporary Bulgarians from the other Balkan nations is the unique signature of SM-Alan people, who appear amongst the direct precursors of contemporary Bulgarians.

References

bioRxiv preprint doi: https://doi.org/10.1101/687384; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1. Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409-13.
2. Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015;522(7555):207-11
3. Mathieson I, Alpaslan-Roodenberg S, Posth C, et al. The genomic history of southeastern Europe. Nature. 2018;555(7695):197-203
4. P. Damgaard et al , 137 ancient human genomes from across the Eurasian steppes, , Nature, Nature Springer, May 9 2018
5. Yunusbayev et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol Biol Evol. 2012;29(1):359–365
6. Sena Karachanak,Viola Grugni, Simona Fornarino, Desislava Nesheva, Nadia Al-Zahery, Vincenza Battaglia, Valeria Carossa, Yordan Yordanov, Antonio Torroni, Angel S. Galabov, Draga Toncheva, and Ornella Semino. Y-Chromosome Diversity in Modern Bulgarians: New Clues about Their Ancestry