

# BALKAN ARCHAEOLOGY, GENEALOGY, AND POPULATION ORIGIN FROM THE PERSPECTIVES OF CULTURAL GENOMICS

*Lolita Nikolova\**

\* Open Global Research Academy, Salt Lake City, Utah, USA; [lnikol@iianthropology.org](mailto:lnikol@iianthropology.org)

**Rezumat.** Articolul de față are scopul de a propune un cadru pentru cercetarea originii populației din Balcani pe baza teoriei fundamentate din perspectiva genomicii culturale.

Genomica culturală este o nouă disciplină complexă care se ocupa cu studierea populației și a originii indivizilor, care înglobează genetica, arheologia, genealogia și alte domenii conexe. Obiectivul său este analiza critică a datelor genetice din perspectiva contextelor culturale diacronice, cu scopul de a releva limitarea și delimitarea inerente explorării originii genetice.

Combi-nația de date genetice, genealogice și arheologice oferă o bază relativ cuprinzătoare pentru o abordare științifică a problematicilor legate de originea populației în ansamblul ei și a originii individuale în Balcani. Balcanii pot fi împărțiți în trei regiuni cultural-istorice care se suprapun: Balcanii de Vest, Balcanii de Sud și Balcanii de Nord. În prezent, regiunea Balcanilor de Vest oferă cele mai ample origini genealogice ale populației locale mergând în trecut pana în secolul al XVI-lea. Datele genetice corelează numeroasele origini individuale occidentale și est-balcanice și demonstrează dinamica demografică extrem de complexă a migrațiilor indivizilor și familiilor.

Datele arheologice furnizează științei informații referitoare la contextul cultural și probele ADN din trecut. Acestea din urmă sunt generale și nu oferă concluzii foarte specifice, cu toate că, în combinație cu date genealogice și testele ADN efectuate asupra populației de viață, ele extinde ipoteza dinamicii demografice ale populației din Balcani, istoric merge înapoi la cele mai timpurii comunități preistorice pentru care există date genetice disponibile. Metodologia prospectiv-retrospectivă permite datelor complexe ulterioare să fie utilizate ca o matrice de control pentru probele anterioare ADN-ului vechi.

Lucrările evaluează în mod critic valoarea științifică a informațiilor din testele ADN comerciale și de cercetare genetică și oferă o metodologie de utilizare critică a acestor dovezi în scopuri științifice.

**Cuvinte cheie:** genomica culturală, populația balcanică, arheologie, genealogie, genomică, genetică, haplogroup T1a1aL.

## 1. Introduction

The study of population origin is one of the traditional topics in the social sciences. The record base has been increasing with the addition of new sources and the research questions also often change. In the past, the main research question was about

ethnogenesis. In Bulgaria, this question was specified as a development of the Bulgarian ethnicity with the participation of three ethnos components—Thracians, Slavic, and Proto(Early)-Bulgarians. Archaeology supported the theory by revealing rich Thracian culture, the genesis of which was a sub-research question. The dominant academic view has been based on grounded theory, which looks at ethnogenesis as a dynamic reproduction that also includes ethnical interactions and migrations.

The blooming field of genomic studies provided new data about the very deep roots of the Balkan population. However, the genomic evolutionary hypotheses of the populations are based on compressed models and, for the time being, they cannot be accepted as truly scientific models in all their parts, especially having in mind that they come from grant reports, not from real scientific works with research questions, a research design, literature review, a detailed analysis of the data and conclusions which can be verified by the readers-experts. Every scientist who uses genomic data or models needs today to critically analyze the methodology, theory, and database as well as the analyses and conclusions of the geneticists. The geneticists are typically not well educated in archaeology and anthropology<sup>1</sup> and deal with different, unrelated scientific problems; furthermore, their conclusions are based on software programs with many limitations. Most importantly, in many cases they repeat even old outdated theories of archaeologists, without cited them (e.g. the works by Maria GIMBUTAS).

This situation of the current approach to population origin is one of the reasons to develop the new discipline of cultural genomics, which attempts to combine numerous evidence from different disciplines to scientifically approach the problems and move toward successful research<sup>2</sup>. Cultural genomics is an open field of research, which embeds innovative emerging methodologies and theories and assists the scholars in searching for the origin, in particular, since genomics has been changing the dynamics of development of the framework for research of the cultural identity of everybody in the world.

One of the most essential elements for the field of cultural genomics is the increased interest in genetic genealogy and the millions of data gathered by commercial companies like Ancestry.com<sup>3</sup>, 23&Me<sup>4</sup>, and FamilyTreeDNA<sup>5</sup>. Important online websites like haplogroup.org<sup>6</sup> and gedmatch.org<sup>7</sup> assist in analyzing and understanding the data. In fact, every scientist can test their own DNA with one or more of the companies and use the data for going deeply insight in the own personal origin, which is also the origin of the lineage and macro-ethnical group to which has belonged.

---

<sup>1</sup> Nikolova, 2018a.

<sup>2</sup> Nikolova, 2018a.

<sup>3</sup> Ancestry, 2018.

<sup>4</sup> 23&Me, 2018.

<sup>5</sup> FamilyTreeDNA, 2018.

<sup>6</sup> haplogroup.org, 2018.

<sup>7</sup> gedmatch.org, 2018.

## 2. Approaches to population origin in the Balkans: reviews of recent selected literature and sources

The first scientific work on cultural genomics as a field in which archaeology, genealogy, and genetics/genomics intersect does not research the problem of population origin in the Balkans<sup>8</sup>.

The archaeological and historical literature on population origin in the Balkans is abundant and concerns mostly the main large periods—the emergence of humans in the Balkans (Paleolithic)<sup>9</sup>, the emergence of the Neolithic population (Neolithic)<sup>10</sup>, the emergence of the Bronze Age<sup>11</sup>, the ethnical migrations in Late Antiquity<sup>12</sup>, and the Middle Age population origin<sup>13</sup>. Depending on the research questions, the arguments are either more general or more detailed.

In this part, we selected the most recent archaeological and genetic works related to problems of Balkan population origin in the Eurasian context to develop a framework of the recent approaches to the problems. The criteria of selection was determining how relevant the publication is to the topic and how it capitalizes on the opportunity to exemplify a specific general direction of current research.

### 2.1. Eurasian context

The problem of earliest Balkan population origin is related to Neanderthals and Cro-Magnon dispersal in Eurasia. Recently, archaeologist Marcel OTTE defended the theory that “*archaeology helps us to recognize the homogeneity of the central Asiatic cultural traditions (‘upper Paleolithic’) and the homogeneity of the anatomy all across this huge area*”<sup>14</sup>. The main scientific argument of the author about the Upper Paleolithic as the period when the big map of the proto-Indo-Europeans (IE) was created is that the archaeological map shows “*the unity between south-west Asia and Europe at the origin of the movement, at least over the past forty-three thousand years ... , which is precisely the geographic area of dispersion of Indo-European populations*”<sup>15</sup>. It can be added that the other two theories (Neolithic and Bronze Age) face the biggest problem: the IE speakers cannot be archaeologically connected to cover the whole territory where the IE languages were distributed. M. Otte sees the main factors for creating language and cultural unity in the later Paleolithic in movement strategies of the modern people in the context of plain environment, long-distance connections, and novelty.

Further, M. Otte states that Neolithic migration from Anatolia has a limited character and its triggers were the non-IE population; this migration was stopped by the Mesolithic population at the Danube, while the acculturation was the main factor in the development of the European pottery cultures in the earlier Neolithic. Unfortunately,

---

<sup>8</sup> Nikolova, 2018a.

<sup>9</sup> Mihailović, 2016; Otte, 2017.

<sup>10</sup> Nikolova, 2007.

<sup>11</sup> Nikolova, 1999.

<sup>12</sup> E.g., Ninić, 1989.

<sup>13</sup> E.g., Fine, 1994.

<sup>14</sup> Otte, 2017, p. 43.

<sup>15</sup> Otte, 2017, p. 48.

the author does not analyze in detail all the further prehistoric periods of Eurasia mentioning the Scythes as the latest IE wave.

In the work of M. Otte, the Upper Paleolithic unity is an extremely important scientific argument for tracing proto-IE communities in this period. Another very strong argument is the reminder that the earliest Anatolian culture was non-IE, based on existing non-IE early records. However, the author did not go so far as to embrace the contact theory of distribution of the languages, which is the strongest explanation of IE language distribution. Also, it is still missing his argument how his theory relates to the genetic phylotrees.

Three groups of archaeological records can be defined as the database for studying population origin: burial data, material culture, and social and natural environment as a living context. Typically, archaeologists study the burial ritual as a marker of population dynamics, complemented by anthropological and genetic data. Recently, many new archaeological data about ancient population origin were simply submitted to the geneticists without detailed independent publication<sup>16</sup>, while the included supplementary information is heterogenous, lacking unified criteria of publication of the archaeological data. At the same time, the geneticists made conclusions without real arguments based on grounded theory and offered generalized title statements behind which exist controversial notes for future serious research. Even more, without expert prehistorians on the writing team and by failing to use qualified expert publications, the archaeological information has been deformed to reflect incorrect chronology and/or statements—the leading marker of archaeological expertise. For instance, it is stated that Pit Grave Culture was distributed into central and western Europe in abt. 2500 BC<sup>17</sup>. Also, the article stated that although “*we find sporadic steppe-related ancestry in Balkan Copper and Bronze Age individuals, this ancestry is rare until the late Bronze Age*”<sup>18</sup>. However, the abstract of the same article refers to “*the early stages in the formation of Bronze Age steppe ancestry*”<sup>19</sup>. Most importantly, although it is a publication produced jointly with archaeologists, the research paper does not analyze the gathered archaeological data for genetic needs, but only discusses general problems referring to literature and makes conclusions in fact without showing a specific record analysis of the software analyses used. So, the whole heterogenous character of the grant report leaves the impression of notes for real scholarly work despite the title promising synthesis and a complete theory.

Peter de Barros DAMGAARD with a team<sup>20</sup> added a new archaeological argument in the discussion of the distribution of the IE languages in Asia, stating that not the Pit Grave Culture but different communities introduced the languages in central Asia, based on the archaeological argument of independent development of horse stockbreeding. This is just one example to show that the grant-based scholars began to feel

---

<sup>16</sup> E.g. Mathieson *et alii*, 2018b.

<sup>17</sup> Mathieson *et alii*, 2018a.

<sup>18</sup> Mathieson *et alii*, p. 5.

<sup>19</sup> Mathieson *et alii*, p. 1.

<sup>20</sup> Damgaard *et alii*, 2018.

the very fragile character of the Pit Grave Culture steppe invasion theory and started to search for new pathways to maintain their access to grants for research. However, it is encouraging that a new generation of serious geneticists will begin to have access to grants for novel and innovative scientific results.

As a matter of fact, the strong statements regarding steppe invasions among geneticists just repeat the well-known, outdated, and much-criticized theory of M. Gimbutas, which in the later 20<sup>th</sup> century was a big scholarly burden<sup>21</sup>. Because of the hybrid character of most recent publications and the combination of texts by different authors, we see that there is in fact no theory, since theory requires non-conflicting inside arguments. One very good example is the work by Morten E. ALLENTOFT with a team<sup>22</sup>. While in the text of the article the authors illustrate the well-known theory of M. Gimbutas trying to represent everything as something new, they include a statement in the supplemental information:

*“The Tripolje populations expanded into the steppe where they encountered Maikop groups and adopted individual burials under barrows and metallurgy. Horse domestication and the development of wheeled vehicles, in the style of later prairie wagons, took place to support a mobile pastoral lifestyle in the steppe. A new expansionist pastoral society, called the Yamnaya Culture [in English Pit Grave Culture] ... evolved from this merging of cultures, and spread rapidly east and westwards, from Hungary to the Ural Mountains. This Yamnaya expansion has often been associated with the spread of Proto-Indo-European languages, after the early split of Anatolian/Hittite”<sup>23</sup>.*

The scientific question is: Why do the same authors believe that in one region the Late Copper Age culture was transformed and the population interacted, while in the east of Europe the Pit Grave Culture, like a tornado, destroyed all cultures that were thousands of years old? Where can we find the population resource for such a mass invasion going to the west, even in the British Islands? In other words, the critical approach to the geneticists’ recent research shows that although they attempted to find genetic data to illustrate the theory of M. Gimbutas, they have very serious problems—even simply logical ones—in what they offer as geneticists’ research and results. Such works have also increased the role of Balkan research.

## **2.2. Genetic arguments**

The genetic literature on population origin in the Balkans is typically at a macrolevel and often lacks significant specifications. The current themes are mainly Y-DNA, mt-DNA, and autosomal data analyses using contemporary DNA samples for population genetics<sup>24</sup>, ancient DNA samples<sup>25</sup>, DNA samples for population genetics, and DNA tests for genealogical needs<sup>26</sup>, as well as a combination of contemporary DNA

---

<sup>21</sup> See Nikolova, 1999.

<sup>22</sup> Allentoft *et alii*, 2015a, 2015b.

<sup>23</sup> Allentoft *et alii*, 2015b.

<sup>24</sup> Karachanak *et alii*, 2012, Делев, 2017.

<sup>25</sup> Mathieson *et alii*, 2018a, 2018b.

<sup>26</sup> Yankova, 2015.

data for population genetics, DNA tests for genealogical needs, and ancient DNA data<sup>27</sup>.

The methodology and software used by the different teams is similar, with diverse implementation of the numeric data in the demographic analyses. Most critical are the research questions, which demonstrate the limitation of the geneticists in the field of demographic research without using the archaeological, non-biological, and genealogical data as integrative or complementary or as a control database<sup>28</sup>.

The general statement is similar to the other regions of Europe: the Balkan population is of western Eurasian origin with insignificant evidence for a population presence whose origin was in eastern Eurasia<sup>29</sup>. Theoretically, it is very problematic, without researching in-depth specific lineages, while the term “gene pool” is substituted to reflect that specific Balkan lineages originated in eastern Eurasia and Africa<sup>30</sup>. However, there is a big difference between genealogical origin and genetic match. In many cases, genetic matches only show very distant common origin, which does not relate to the specific recent lineage documented.

The general conclusions from the study of contemporary DNA samples are still only a preliminary step toward in-depth research of population origin in the Balkans. The Bulgarian data still do not have regional articulation and diachronic explanation, although Y-DNA and mt-DNA haplogroups of contemporary individuals (sample size <800>900) have been predicted<sup>31</sup>. The well-documented Y-DNA haplogroups in the studied samples were E1b1b1a, I1, I2a, J1, J2, R1a, R1b, and G2a. Very informative, however, are the data from the Bulgarian DNA Project at FamilyTreeDNA<sup>32</sup>, where the researchers can find not only the predicted/confirmed haplogroup, but also information about the place of origin, the SNIPs, the ancestral origin, and even pedigrees. For instance, E1b1b1a has a subclade E1b1b1a1b that is well documented in Bulgaria. There is also E1b1b1c (M123), the distribution of which is in southeastern Bulgaria, northwestern Bulgaria, and a single example from north Greece (Bulgarian name). Other documented Y-DNA for genealogical needs are as follows: G1a, G2a1a, G2a2a, G2a2b1, G2a2b2 (8 samples), G2b2, H1a, I1a1b, I1a2, I1a3, I2a1b2a1, I2a1b2a1a, I2a1b2a1c, I2a2a, I2a2b, I2c, J1, J2a1\*, J2a1b, J2a1h, J2a1h2a1, J2a1h2c, J2a2, J2b\*, J2b1, J2b2, L2, Q, R1a1a, R1a1a1b1a1a, R1a1a1b1a1b1, R1a1a1b1a2a, R1a1a1b1a2b, R1a1a1b2, R1b1a1a2\*, R1b1a1b2a1a, R1b1a1b2a1a2a, R1b1a1b2a1a2b, R1b1a1b2a2, R1b1a1b2a2b, R1b1a1b2a2c1a, T1. These data provide important details of the Y-DNA of the contemporary Bulgarian population, since it is correlated with maps that show the distribution of the haplogroups by specific places. The spreading of the different haplogroups over most of the regions of Bulgaria is due to the inner migration of members of the lineages. Without genealogical information, such dynamics of the population are explained by the absence of genetic difference between the ethnographic regions

---

<sup>27</sup> Nikolova, 2018a.

<sup>28</sup> Nikolova, 2018c.

<sup>29</sup> Karachanak *et alii*, 2012; Kovacevic *et alii*, 2014.

<sup>30</sup> Kovacevic *et alii*, 2014, p. 2.

<sup>31</sup> Karachanak *et alii*, 2012; 2013; Karachanak-Yankova *et alii*, 2015.

<sup>32</sup> Delev, 2018.

of Bulgaria<sup>33</sup>, although the genetic data themselves do not have ethnical information without a detailed contextual analysis. Obviously, the research question for determining the difference is not which haplogroups are distributed in the different parts of Bulgaria, but how the different lineages interrelate and which lineages do not relate at all. In other words, comparing the geneticists' studies with genealogical genetic studies shows that the geneticists' studies benefit only if they use the available genealogical genetic information.

This is very well demonstrated by comparing the Balkan genetic genealogical data. The following Y-DNA haplogroups have been documented in the Balkan genetic group at FamilyTreeDNA [(sample size [n]=350): E1b, E1b-V13, G1, G2a, H, I1, I2a, J1, J2a, J2b, N, R1a, R1b, T1a2<sup>34</sup>].

With such a reference, we can compare the regional studies of all Balkan countries. For instance, the research of the Y-DNA of Croatian men inferred that haplogroups I, R1a1, and R1b are most common, followed by E3b, J2a, G, L, and T<sup>35</sup>. Some authors even stated that the populations of western Balkan countries share a large fraction of the ancient gene pool of Southeastern Europe, where 70% of the paternal lineages consist of five European-specific Y-DNA haplogroups: E3b1, I-P37(xM26), J2, R1a, and R1b<sup>36</sup>. However, in this and all other cases, including the FamilyTreeDNA groups, the size of the samples and the chosen methodology may have considerable impact on the conclusions.

The mt-DNA of Balkan Genetic group at FamilyTreeDNA includes the following haplogroups (n=152): H, HV, I, J, K, N, T, U, V, W, and X<sup>37</sup>. The Bulgarian sample of <900>1000 resulted in the prediction of the following mt-DNA haplogroups: H, H5, HV, HV0, R0a, J, U1, U2, U2e, U3, U4, U5a, U5b, U6, U7, U8, K, N1, N2, X, M, T1, T2<sup>38</sup>.

The contemporary research of Balkan population genetics at the macrolevel shows that regional population differences cannot be distinguished based on the haplogroups because of the genetic unification of the region. However, the genealogical genetic data demonstrate the way the lineages migrate and create dynamic models of the development of cultural identity. Then, closer interaction between population genetics and genetic genealogy will create models of more articulated scholarly hypotheses about the origin of the Balkan population during different periods from the beginning of human culture in the region until today.

The ancient data, for the time being, are not numerous and cannot build frameworks of ancient genealogical trees<sup>39</sup>.

---

<sup>33</sup> Karachanak *et alii*, 2012.

<sup>34</sup> Inal, 2018.

<sup>35</sup> Mršić *et alii*, 2012.

<sup>36</sup> Kovacevic *et alii*, 2014 and ref. cited there.

<sup>37</sup> Inal, 2018.

<sup>38</sup> Karachanak *et alii*, 2012.

<sup>39</sup> Hешева, 2016; Mathieson *et alii*, 2018.

### 2.3. Compressed genetic models

The first variant is the prehistoric model of origin based on the haplogroups. The current branch of researchers follows a specific direction of origin, evolution, and distribution of the different haplogroups. One of the extremes is that, when writing about origin, the British Neolithic population would be anchored directly to the New East, since geneticists do not ask the right question: What is the origin of the British population? Instead, they ask: What is the origin of the European population? This creates an opportunity for hyper-macrolevel analysis, far removed from the real scientific goals of the social research.

Another extreme is the hyper-macrolevel of the steppe origin of the British Bronze population, which again is based on abstract models and lacks real arguments since only selected data have been used<sup>40</sup>.

In the first case (Near East-British Islands model), it will be wrong to state that the origin of the British Neolithic population was in the Balkans just because the Neolithic population settled first in the Balkans. Also, some archaeologists do not even accept that the Neolithic immigrants reached the British Islands (see above), while the genetic model is based on a specific macro-haplogroup presumably derived in Europe but with Near East origin. Accordingly, geneticists first need to prove that there was a crucial Neolithic migration based on archaeological data (positive and negative arguments) and then discuss the genetic data, which are very vague and can be interpreted from many perspectives. The current state of research shows that the Holocene was a period that stimulated independent sedentarization while the interaction between lineages, including intermarriages and movement, were typical of Eurasia from the beginning of human history<sup>41</sup>. The genetic autosomal data provide interconnected similarity between the haplogroup at a different scale, depending on the quantity of the shared centiMorgans. However, they do not point to an exact place or a very precise time when the event happened or where the most recent common ancestors occurred. The compressed mass autosomal data are like the cognitive map of the people—they need to be placed in the diachronic matrix and analyzed in detail.

Regarding the second case (Russian Steppes-British Islands model), it is archaeologically well documented that even in the Balkans (which is much closer to the Russian steppes) the steppe population did not play the role of mass invaders during the Final Copper and Early Bronze Age. Smaller groups are archaeologically documented at different scales, and they penetrated different parts of the Balkans and interacted with the local populations. Their cultural behavior is that of refugees rather than invaders. Again, as in the case of the Neolithic genetic compressed models, limited data have been analyzed from the perspectives of an already created model without using the method of control data.

The compressed character of the models of ancient origin of the current population is also well documented in all of the so-called ethnical maps coming from the commercial companies for genealogical needs. It was even documented that in determining

---

<sup>40</sup> Nikolova, 2018a and ref. cited there.

<sup>41</sup> Otte, 2017.



the given individual origin, a company may not use the new data<sup>42</sup>. Since none of the companies explains the methodology of how they determine ancient origin, it is difficult to analyze the differences. Which part is due to methodology, to theory, to the programs, to the software, and/or to the lab workers?

### **3. Cultural genomics and the problems of population origin in the Balkans**

Cultural genomics is a new complex discipline for studying population and individual origin that integrates genetics, archaeology, genealogy, and other related fields<sup>43</sup>. Its goal is to critically analyze genetic data through the lens of diachronic cultural contexts with the objective of revealing the limitation and delimitation inherent in the exploration of genetic origin.

#### **3.1. A complex model of population origin in the Balkan**

The complex models of population origin in the Balkans include all available data that provide positive and negative arguments for a given theory—archaeological, historical, genetic, paleogeographic, linguistic, and any other piece of data with origin information. The complex models are open since the limitation of all records to study origin is enormous and it is impossible to formulate a conclusion that does not require further research.

The current state of development of genomic studies on population origin requires using the genomic data only as a source for critical analysis in cultural theories, along with all other possible data. Complex grounded theory can be proposed that includes the following interactive levels of research:

- 1.** Formulation of the research question and sub-research questions. This is a critically important part of the research. Many of the geneticists' works do not have strong conclusions because their research question(s) does not guide them toward deep research of the problems.
- 2.** Development of a framework of research (or research design).
- 3.** Literature and source review. This is a very important stage that geneticists typically miss in their publications, especially when they discuss origin without reference to the archaeological literature on the problems.
- 4.** Gathering of new data for complex analysis and/or development of a new method. For example, in this research, the data of population genetics studies were tested against and compared with the genealogical genetic studies, which created a new scholarly framework already in the literature and source review.
- 5.** Clearly demonstrating the gaps in the previous research based on the literature review and the gathered database.
- 6.** Providing conclusions that answer the research question and direct others toward further research to fill the newly revealed gaps in the research.

This model requires an integration of experts in different fields (archaeology, genetics, genealogy, history, linguistics, in particular).

---

<sup>42</sup> Nikolova, 2018b.

<sup>43</sup> Nikolova, 2018a.

### 3.2. Case study: mt-DNA T1a1L haplogroup

As a starting point, we need to keep in mind that the genealogical models of population origin in the Balkans based on the current DNA tests show a very dynamic picture of the lineages. This is very well documented in the data from FamilyTreeDNA for my case study (haplogroup T1a1L)<sup>44</sup>. Three maps from FamilyTreeDNA<sup>45</sup> show the mechanism of the software—from a detailed determination of the migration of the members of lineages toward a general compressed model, which in fact reflects thousands of years of multidirectional movement of the lineages. Typically, the last map can be seen in different variants in the genetic studies measuring genetic distances. However, every single piece of data has its own genealogy at a given chronological span, while the software flattens everything. This gives reason to propose that a valid scientific model of interpretation of the archaeological, anthropological, and genetic data about population origin is multidirectional lineage interaction, while every migration model based on archaeological, anthropological, and genetic data requires an extremely detailed, complex analysis of all available positive and negative arguments.

The case study with my haplogroup (T1a1L) was analyzed from the perspectives of genealogy and archaeology elsewhere<sup>46</sup>. In this work, I will focus on the autosomal similarities  $> 20 < 9$  centiMorgans, including the data that correlate with mt-DNA.

One of the important programs for comparing genetic data is gedmatch.com. It provides the opportunity to compare and find autosomal genetic matches based on a small number of shared centiMorgans. In the case of my maternal genealogy, the DNA raw data were compared with the data in the entire database of gedmatch.com and showed 2,000 matches with other individuals, many of whom also have entered mt-DNA and/or Y-DNA information of the haplogroup. Our method attempted to correlate the matches according to the haplogroup and to reveal the crossing haplogroup(s), with predicted most recent common ancestors according (MRCA) to the program used. We also attempted to compare not only the number of shared centiMorgans but also the character of the genetic match—the chromosome and the SNP (one-to-one match, V2.1.1(c)), but these results will be reported elsewhere.

**Tables 1–4** show that that there is a considerable diversity, although some patterns can be revealed:

$>20 < 16$  cM (Table 1). Subclades of mt-DNA H haplogroup dominate, although there is one case of K1a1b and two cases of T (T1 and T2b). Most interestingly, the mt-DNA has been correlated with subclades of a very ancient Y-DNA haplogroup (I and E), and R only as an exception (one case).

$>16 < 13$  cM (Table 2). Again, subclades of mt-DNA H haplogroup dominate, complemented by subclades of U, L, T, I, and K. The mt-DNA data correlate mainly with subclades of Y-DNA haplogroups R and I.

---

<sup>44</sup> Nikolova, 2018a; Nikolova, 2018b.

<sup>45</sup> Nikolova, 2018b.

<sup>46</sup> Nikolova, 2018a; Nikolova, 2018b.

>13≤10 cM (Table 3). In this group subclades of mt-DNA H haplogroup continue to dominate, followed by subclades of mt-DNA U and V haplogroups, and of mt-DNA T haplogroup. Single cases of subclades of V, D, K, W, X and J mt-DNA expand the knowledge on the deep genetic interconnection of this case stud of T1a1L mt-DNA haplogroup, which most recent origin was in Bulgaria with roots in Macedonia.

>10≤9 cM (Table 4). Subclades of mt-DNA U began to dominate, followed by H, K, T, A, R, J, X and single cases of V, W and D.

The open research question is whether the predicted by gedmatch.org generations to MCRA (most common recent ancestors) are real. In case of genetic match of T1a1L with H2a2a2 the existed pedigree is rooted in England / Scotland. The gedmatch.com prediction is 5.2 generations. However, the diversity of the genetic matches points to much deeper common genetic roots. Both, 23&Me<sup>47</sup> and National Geographic<sup>48</sup> projects for instance, documented Neanderthal ancestry in the DNA samples – 1.5% or 233 Neanderthal variants (n=1436, n-number of the tested markers). Also, as the results show above, the researched T1a1L mt-DNA case study indeed shares common centiMorgans with subclades of U mt-DNA haplogroup considered as the most ancient in Europe<sup>49</sup>, although there is no a methodology offered by geneticists to build a chronology of the individual samples.

The case study shows, that the claimed by the geneticists absence of diversity in the Balkan and European population is due to the compressed model of the used software and the fact that the number of the haplogroups is very small in compare to the subclades and genotypes. Thanks to genealogy and the advanced in the genetic genealogy globally, the Balkan population map will be supplied by articulated meaningful population diversity in distant future. It also demonstrates the richness of the genetic data if they have been analyzed in depth and in details.

How does archaeology relate to population origin based on DNA individual data? For the time being, the cultural identity of the individuals includes the values of the place of residence and its history complimented by the places of origin and eventually former residence / place of birth (if different). The future development of genetic studies will also add the value of the deep origin of the lineage and of the bigger ethnical group, and possible connection of the lineage with specific archaeological cultures.

#### **4. Theory and conclusion**

Archaeological data supply the cultural genomics with a cultural context and ancient DNA evidence. The latter are general and do not provide very specific conclusions, although in combination with genealogical data and the DNA tests conducted on the living population, they expand the hypothesis of the demographic dynamics of the Balkan population historically going back to the earliest prehistoric communities for which genetic data is available. The prospective-retrospective methodology allows

---

<sup>47</sup> 23&Me, 2018.

<sup>48</sup> National Geographic, 2018.

<sup>49</sup> Cokoş, 2017, p. 4.

the later complex data to be used as a control matrix for the earlier ancient DNA evidence.

The genealogical research on population origin in the Balkans includes numerous lineages, the pedigrees of which may go back even into the Middle Ages. Unfortunately, this literature, based mainly on the traditional genealogy, is not systematized, although mapping the different studied lineages in the Balkans will create comprehensive models of the dynamics and evolution of the Balkan lineages, which typically intersect through traditionally used terms like Bulgarians, Romanians, Serbs, Albanians, Slavs etc.

The genetic ancient and contemporary data have been increasing, but macro-theories and compressed models still dominate. This work showed that every DNA sample needs a most detailed analysis in order to create a framework for general conclusions. Cultural genomics offers slow but steady progress through a complex research and detailed analysis of data not only from genetics, but also from all related disciplines. The research is extremely very time-consuming, although the progress in the integration of scholars from different branches of social sciences and genomics will create a revolutionary scholarly context for significant social scientific results and will open new perspectives for innovative meaningful research.

**Tables 1. Genetic mt-DNA and Y-DNA matches of T1a1L haplogroup case study (>20<16 cM).**

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
<b>H1</b>	I2a2	18.3	18.3
<b>H1</b>	I2a2	18.3	18.3
<b>H</b>		18.1	11.8
<b>T1</b>		17.9	12.3
	E-V13	17.9	9.8
<b>H</b>	E1B1A	17.7	11.2
<b>T2b</b>	I-P37	17.1	9.0
<b>K1a1b</b>	E-Z16661	17	11.1
<b>H1</b>	R-M405	16.7	16.7

**Tables 2. Genetic mt-DNA and Y-DNA matches of T1a1L haplogroup case study (>16<13 cM).**

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
<b>U4a2</b>	J2b2*	15.7	10.6
<b>H13a1d</b>	R-PF7562	15	15.0
<b>U5a1</b>	I1	14.9	7.9
<b>T1a</b>		14.8	8.6
	R1a	14.6	8.1
<b>A8</b>	R1b1b2a1a2d	14.3	7.6
<b>H2a2b</b>		14.2	8.4
<b>H15a</b>		14.1	7.6
<b>K1a-C150T</b>	I2-PH908	14	8.5
<b>H2a1a</b>		13.9	7.5
<b>I3a</b>		13.9	7.9
<b>I-L233 'western'</b>	I-P37	13.8	7.3
<b>I-L233 'western'</b>	I_P37	13.8	7.3
<b>H</b>		13.7	7.1
	I-P37	13.7	13.7
	RM-269	13.6	7.6
<b>U5</b>		13.6	7.0
<b>H1c</b>		13.5	7.0
<b>H1c</b>		13.5	7.0
<b>H1c</b>		13.5	7.0
<b>H8c1</b>		13.5	8.2
	R1b1a2a1a2b1a1	13.4	7.9
<b>H1g</b>	I2b1*	13.3	8.3
<b>T1a1</b>	R-M198	13.3	7.3
<b>H1</b>		13.3	13.3
<b>H1</b>	I1	13.3	13.3

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
<b>U4</b>	R-M417	13.3	7.0
<b>H1</b>		13.2	7.0
<b>U4a2</b>	R1b1b2a1a2f*	13.1	13.1
<b>H12</b>	E1b1b1a2*	13.1	8.0
<b>H5a1</b>	R1b1b2a	13.1	13.1
<b>L2d</b>		13	13.0
<b>Ua1a1</b>	R-L2	13	7.4
<b>U4a2</b>	I1	13	7.2

**Tables 3. Genetic mt-DNA and Y-DNA matches of T1a1L haplogroup case study (>13≤10 cM).**

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
<b>I5A4</b>	R-L260	12.8	7.4
<b>T2b5</b>		12.8	7.2
	G2a3b1	12.8	7.1
<b>U5b3a</b>	R1a1a	12.7	12.7
<b>U5b3a</b>	R1a1a	12.7	12.7
	R-M269	12.7	7.4
<b>H1</b>		12.5	12.5
<b>I2</b>		12.4	7.0
<b>T2B2b1</b>		12.4	7.4
<b>H5a1f</b>	R-M269(R1b1a2)	12.2	7.1
<b>T2c1</b>		12.2	7.0
<b>H1a1</b>		12.2	12.2
<b>U5a1a1</b>		12	12.0
<b>H</b>		11.9	11.9
<b>V7a</b>		11.7	11.7
<b>U5b1e</b>	R1b1b2a	11.6	11.6
<b>D5a2d1</b>		11.4	11.4
<b>U4a2</b>	J2b2*	11.3	11.3
<b>H1a1</b>	R-M198	11.2	11.2
<b>U5a1a1</b>		11	11.0
<b>U4a2</b>	R1b1b2a1a2f*	10.9	10.9
<b>H</b>	R-M417	10.9	10.9
<b>H5</b>		10.8	10.8
<b>W3a</b>	E1b1b1a2*	10.8	10.8
<b>H2a1</b>	R1b1b2a1a2f*	10.8	10.8
<b>H5b</b>	E1b1b1a2*	10.8	10.8
<b>H5b</b>		10.8	10.8
<b>H31</b>	R-M269	10.8	10.8
<b>V</b>		10.7	10.7

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
H	I-M 253	10.7	10.7
V		10.6	10.6
V		10.6	10.6
V		10.6	10.6
U4a2	J2b2*	10.6	10.6
J2b1a	I2a2	10.6	10.6
HV1a2	I2a2	10.6	10.6
U6a7	R-L51	10.5	10.5
H5	R1a1a	10.5	10.5
H4a2	I-M223	10.4	10.4
H		10.3	10.3
H		10.2	10.2
H		10.1	10.1
X2f		10.1	10.1
H5	L1	10.1	10.1
H2a2a2	G2a2b1	10.1	10.1
K1a-C150T		10	10.0

**Tables 4. Genetic mt-DNA and Y-DNA matches of T1a1L haplogroup case study (>10≤9 cM).**

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
	T1a1 1454	9.9	9.9
H	I1	9.9	9.9
H1e1a		9.9	9.9
R0	I2-PH908	9.9	9.9
K1a1	I2a2	9.9	9.9
J1c5a		9.9	9.9
A		9.8	9.8
U8a1a		9.8	9.8
	J2b1	9.8	9.8
H1b	R1a1a	9.8	9.8
H1	R-L51	9.8	9.8
U5b2a2		9.7	9.7
U3a1	R1b1b2a1a2d3*	9.7	9.7
U5b2a1		9.7	9.7
V3	R1b1b2a1a2f*	9.7	9.7
X2f		9.6	9.6
K1c2		9.5	9.5
	I2a2 - M427	9.5	9.5
A2e	I-BY13534	9.5	9.5
T2b	R1b1b2a1a2f*	9.4	9.4

<i>Mt-DNA haplogroup</i>	<i>Y-chromosome DNA haplogroup</i>	<i>Total cM</i>	<i>Largest cM</i>
<b>H5a</b>		9.4	9.4
<b>I2A</b>	Y-46	9.4	9.4
<b>H1</b>		9.4	9.4
<b>HV8</b>	R1b1b2a1a1d	9.3	9.3
<b>X2e2a</b>	E-M4254	9.3	9.3
<b>K1C2</b>		9.3	9.3
<b>T1a</b>		9.3	9.3
<b>H1C2</b>	P-109	9.3	9.3
<b>K</b>		9.3	9.3
<b>R0a1a</b>	I2a2	9.3	9.3
<b>U4a1a</b>		9.3	9.3
<b>X2f</b>		9.2	9.2
<b>U5b2a1b</b>		9.2	9.2
<b>H1</b>		9.2	9.2
	R1a-M198	9.2	9.2
<b>T2B</b>	G-M201	9.2	9.2
<b>H3</b>	R1b1b2a1a1d	9.2	9.2
<b>H</b>	R1b1b2a1a1d	9.2	9.2
<b>H11a</b>		9.1	9.1
<b>HV</b>	E1b1b1a2*	9.1	9.1
<b>H</b>	R-Z18	9.1	9.1
<b>H27</b>	I2a1b2a1	9.1	9.1
<b>H</b>	R1b1b2a1a2f*	9.1	9.1
<b>K1a</b>	E1b1b1a2	9.1	9.1
<b>A12a</b>	R-M269	9.1	9.1
<b>H3a</b>	R1a1a	9.1	9.1
<b>J2a1a1b</b>		9.1	9.1
<b>U5a1a1</b>	R1b1b2a1a	9.1	9.1
<b>H3</b>		9.1	9.1
	R1b-L11	9.1	9.1
<b>U5A1</b>	R-M269	9.1	9.1
<b>W3a</b>	R1b1b2a1a2f*	9	9.0
<b>D1</b>		9	9.0
	R-M269	9	9.0
<b>H10a1</b>	R1b1b2a1a	9	9.0
<b>H13a1c</b>		9	9.0
<b>T1a1</b>	R1b1b2a1a1	9	9.0
<b>T2b17a</b>		9	9.0



## BIBLIOGRAPHY

- Делев, 2017      Евгени ДЕЛЕВ, *Българите, основоположници на европейската цивилизация*, БОМИ, София, 2017, 240 p., ISBN 978-619-90190-8-5.
- Нешева, 2016      Десислава В. НЕШЕВА, *Характеризиране генетичните корени на населението по българските земи*, Автореферат на дис. труд за ... степен «доктор», 2016, Медицински университет, София, 72 p.
- 23&Me, 2018      *23&ME*, *23&Me*, 2018, <http://www.23andme.com>
- Allentoft *et alii*, 2015a      Morten E. ALLENTOFT, Martin SIKORA, Karl-Göran SJÖGREN, Simon RASMUSSEN, Morten RASMUSSEN, Jesper STENDERUP, ..., Eske WILLERSLEV, *Population genomics of Bronze Age Eurasia*, In: *Nature*, 522, 2015, p. 167-172.
- Allentoft *et alii*, 2015b      Morten E. ALLENTOFT, Martin SIKORA, Karl-Göran SJÖGREN, Simon RASMUSSEN, Morten RASMUSSEN, Jesper STENDERUP, ..., Eske WILLERSLEV, *Population genomics of Bronze Age Eurasia*, Supplementary Information, *Nature*, 522, 2015, p. 1-46.
- Ancestry, 2018      *Ancestry.com*, 2018, <http://www.ancestry.com>
- Cocoş *et alii*, 2017      Relu COCOŞ, Sorina SCHIPOR, Montserrat HERVELLA, Petro CIANGA, Roxana POPESCU, Claudia BĂNESCU, Mihai CONSTANTINESCU, Alina MARTINESCU, Florina RAICU, *Genetic affinities among the historical provinces of Romania and Central Europe as revealed by an mtDNA analysis*, In: *BNC Genetics*, 18, 20, 2017.
- Damgaard *et alii*, 2018      Peter de Barros DAMGAARD, Rui MARTINIANO, Jack KAMM, J. Victor MORENO-MAYAR, Goos KROONEN, Michaël PEYROT, ... & Eske WILLERSLEV, *The first horse herders and the impact of early Bronze Age steppe expansions into Asia*, In: *Science*, 360, 6396, eaar7711, 2018, DOI: 10.1126/science.aar7711.
- Delev, 2018      Evgeni DELEV, Adm., Bulgarian DNA project, *FamilyTreeDNA*, 2018, <http://www.familytreedna.com>
- FamilyTreeDNA, 2018      *FamilyTreeDNA*, 2018, <http://www.familytreedna.com>
- Fine, 1994      John V. A. FINE, JR., *The late medieval Balkans: A critical survey from the late twelfth century to the Ottoman conquest*, The University of Michigan Press, Ann Arbor, MI, 1994, 683 pg., ISBN 0-472-10079-3.
- gedmatch.org, 2018      *gedmatch.org*, 2018, <http://www.gedmatch.org>
- haplogroup.org, 2018      *haplogroup.org*, 2018, <http://www.haplogroup.org>
- Harvati, Roksandic, 2016      Katerina HARVATI, Mirjana ROKSANDIC (eds), *Paleoanthropology of the Balkans and Anatolia: Human evolution and its context*, Springer, Dordrecht, 2016, 330 pg., ISBN 978-94-024-0873-7.
- Inal, 2018      Erhan INAL (Admin)., *Balkans: Genetic analysis of people from different Balkan countries*, *FamilyTreeDNA group*, 2018, on-line at <http://www.familytreedna.com>

- Karachanak *et alii*, 2012 Sena KARACHANAK, Valeria CAROSSA, Desislava NESHEVA, Anna OLIVIERI, Maria PALA, Baharak Hooshiar KASHANI,... & Antonio TORRONI, *Bulgarians vs the other European populations: a mitochondrial DNA perspective*, In: *International Journal of Legal Medicine*, 126, 4, 2012, p. 497-503.
- Karachanak *et alii*, 2013 Sena KARACHANAK, Viola GRUGNI, Simona FORNARINO, Desislava NESHEVA, Nadia AL-ZAHERY, Vincenza BATTAGLIA, ... & Ornella SEMINO, *Y-Chromosome diversity in modern Bulgarians: New clues about their ancestry*, In: *PLoS ONE*, 8, 3, e56779, 2013, p. 1-9.
- Karachanak-Yankova *et alii*, 2015 Sena KARACHANAK-YANKOVA, Desislava NESHEVA, Angel S. GALABOV,... & Draga TONCHEVA, *Distribution of East Eurasian Y-Chromosome and mitochondrial DNA haplogroups across Eurasia: Insights into the genetic ancestry of Bulgarians*, In: *An Academic Publisher*, 5, 4, 2015, p. [205]-266.
- Kovacevic *et alii*, 2014 Lejla KOVACEVIC, Kristiina TAMBETS, Anne-Mai ILUMÄE, Aleina KUSHNIAREVICH, Bayazit YUNUSBAYEV, Anu SOLNIK, ... & Damir MARJANOVIĆ, *Standing at the Gateway to Europe - The genetic structure of western Balkan populations based on autosomal and haploid markers*, In: *PLoS ONE*, 9(8), e10509, 2014, p. 1-15.
- Mathieson *et alii*, 2018a Iain MATHIESON, Songül ALPASLAN-ROODENBERG, Cosimo POSTH, Anna SZÉCSÉNYI-NAGY, Nadin ROHLAND, Swapan MALLICK, ... & David REICH, *The genomic history of Southeastern Europe*, In: *Nature*, 555, 2018, p. 197-203.
- Mathieson *et alii*, 2018b Iain MATHIESON, Songül ALPASLAN-ROODENBERG, Cosimo POSTH, Anna SZÉCSÉNYI-NAGY, Nadin ROHLAND, Swapan MALLICK, ... & David REICH, *The genomic history of Southeastern Europe*, Supplementary Information, In: *Nature*, 555, 2018, p. 1-58.
- Mihailović, 2014 Dušan MIHAILOVIĆ (ed), *Palaeolithic and Mesolithic research in the central Balkans*, Serbian Archaeological Society, Commission for the Palaeolithic and Mesolithic, Belgrade, 2014, 160 p., ISBN 978-86-913229-9-1.
- Mršić *et alii*, 2012 Gordana MRŠIĆ, Branka GRŠKOVIĆ, Andro VRDOLJAK, Maja POPOVIĆ, Ivica VALPOTIĆ, Šimun ANĐELINOVIĆ, ... & Dragan PRIMORAĆ, *Croatian national reference Y-STR haplotype database*, In: *Molecular Biology Reports*, 39, 2012, p. 7727-7741.
- National Geographic, 2018 NATIONAL GEOGRAPHIC, *National Geographic Geno 2.0.*, 2018, <http://www.geographic.national.com>
- Nikolova, 1999 Lolita NIKOLOVA, *Balkan in later Prehistory*, British Archaeological Reports International Series, 791, Oxford, 1999, 442 p., ISBN 1-84171-108-X.
- Nikolova, 2007 Lolita NIKOLOVA, *Toward an evolutionary model of gradual development of social complexity among the Neolithic pottery communities in the Balkans*, In: Michela SPATARO, Paolo BIAGI (eds), *A short walk through the Balkans: the first farmers of the Carpathian basin and adjacent regions: proceedings of the Conference held at the Institute of Archaeology UCL on June 20th-22nd, 2005*, Quader-

- no, 12, Società per la preistoria e protostoria della regione Friuli-Venezia Giulia, Trieste, 2007, 237 p.; p. 89-102.
- Nikolova, 2018a Lolita NIKOLOVA, *Cultural genomics and the changing dynamics of cultural identity*, Nova Science, New York, NY, 240 p., ISBN 978-1-53614-073-6.
- Nikolova, 2018b Lolita NIKOLOVA, *From “Story for” to “Reference to”: Genetic genealogy and origin setting*, In: *Global Journal of Archaeology and Anthropology*, 6, 5, 2018, p. 1-22.
- Nikolova, 2018c Lolita NIKOLOVA, *Genetic Genealogy as an Emerging Research Branch of Human Ancestry and Origin (Review of Curricula and Other Learning Sources)*, In: *International Journal of Humanities Social Sciences and Education*, 5, 3, 2018, p. 169-183.
- Ninić, 1989 Ivan NINIĆ, *Migrations in Balkan history*, Srpska akademija nauka i umetnosti, Beograd, 1989, 170 p., ISBN 86-7179-006-1.
- Otte, 2017 Marcel OTTE, *Indo-Europeans arrived in Europe with modern man*, In: *Philology*, 3, 2017 (2018), p. 43-56.
- Yaneva, 2015 Mariana YANEVA, *A genetic analysis of the people currently inhabiting the country of Bulgaria*, In: *Research Gate*, 2015, p. 1-24.