4-4-2017

Landscape Complexity in the Caucasus impedes Genetic Assimilation of Human Populations More Effectively than Language or Ethnicity

David Tarkhnishvili
Ilia State University, Tbilisi, david_tarkhnishvili@iliauni.edu.ge

Alexander Gavashelishvili
Ilia State University

Marine Murtskhvaladze
Ilia State University, Tbilisi

Ardashel Latsuzbaia
Ilia State University, Tbilisi

Landscape Complexity in the Caucasus impedes Genetic Assimilation of Human Populations More Effectively than Language or Ethnicity

David Tarkhnishvili ¶*[1], Alexander Gavashelishvili[¶1], Marine Murtskhvaladze[1], Ardashel Latsuzbaia[1]

[1]Ilia State University, Cholokashvili Str. 5, 0162 Tbilisi, Georgia

Corresponding author: David Tarkhnishvili, Quantitative Ecology Group, Institute of Ecology, Ilia State University, Cholokashvili Str. 5, 0162 Tbilisi, Georgia

Issue: 88.4

Running header: Autosomal differentiation in Caucasus

Keywords: Y-chromosome DNA markers, autosomal markers, Caucasus, Glacial Refugia, human ecology, landscape genetics, ethnogenesis.

¶These authors contributed equally to this work.

* corresponding author; email: *david_tarkhnishvili@iliauni.edu.ge*

Abstract

The analyses of 15 autosomal and 23 Y-chromosome DNA STR loci in five rural populations from the Caucasus, namely four ethnically Georgian and one ethnically Armenian populations, indicated that two populations of Georgians - one from the west and the other one from the east of the Greater Caucasus Mountains - were both patrilineally and autosomally most differentiated from each other, and the other populations of Georgians and Armenians held an intermediate position between those two. This pattern may be due to human dispersal from two distinct glacial refugia in the last glacial period and the early Holocene, followed by less gene flow among the populations from the Greater Caucasus than those from the rest of the Caucasus, where the populations have undergone substantial admixture in historical time. This hypothesis is supported by a strong correlation between genetic differentiation among the populations and landscape permeability to human migrations. The latter is determined by terrain ruggedness, forest cover, and snow cover. Although geographic patterns of autosomal and Y-chromosome DNA are not fully concordant, both are influenced by landscape permeability, and show a similar east-west gradient. Our results suggest that this permeability was a stronger factor limiting gene flow among human populations in the Caucasus than ethnic or linguistic boundaries.

The human population of the Caucasus, the mountain region between the Black and Caspian Seas, is relatively well studied with molecular genetic methods (Bulayeva et al., 2003; Nasidze et al., 2003, 2004a,b; Marchiani et al., 2008; Balanovsky et al., 2011; Yunusbayev et al., 2012; Tarkhnishvili et al., 2014). People that belong to different ethnic and linguistic groups of the Caucasus are genetically closest to the populations of West Asia and Eastern Mediterranean, although they show certain genetic distinctiveness (Nasidze et al., 2004b; Balanovsky et al., 2011; Busby et al., 2015). There are also substantial genetic differences between the ethnic and subethnic groups of the Caucasus, which often exceed the differences between geographically more distant human populations of Europe (Nasidze et al., 2004b). The local genetic differences follow geographic gradients rather than linguistic disruptions (Marchiani et al., 2008) and genetics may differ between populations of the same ethnic group due to living in different physical environment (Tarkhnishvili et al., 2014).

There is a clear east-west gradient in Y-chromosome DNA haplogroup frequencies in the Caucasus. People with Y-chromosome DNA haplogroup G2a prevail in the west of the Greater Caucasus Mountains regardless of ethnic or linguistic affiliation, whereas people with the haplogroups J1 and J2 dominate in the east of the Greater Caucasus (Balanovsky et al., 2011; Yunusbayev et al., 2012; Tarkhnishvili et al., 2014). A similar east-west gradient is observed in ethnic groups south of the Greater Caucasus, and holds for the entire ethnic range of Georgians (Tarkhnishvili et al., 2014).

This east-west gradient of the frequencies of Y-chromosome DNA haplogroups G and J across the Caucasus is probably inherited from the Paleolithic time. Haplogroup J is almost certainly derived from ancestral populations in the Fertile Crescent (Cinnioğlu et al., 2004; Battaglia et al., 2008; Tofanelli et al., 2009;Grugni et al., 2012; Gavashelishvili and Tarkhnishvili, 2016), where people of this haplogroup probably survived the last glacial period, including the Last Glacial Maximum (LGM). Haplogroup G (which diverged from the other haplogroups ca. 50 kya (Wei et al., 2012; Karmin et al., 2015), most likely persisted during the last glacial period in the Caucasus (Gavashelishvili & Tarkhnishvili, 2016). The

majority of the languages spoken in the northeastern Caucasus, the region where Y-chromosome DNA haplogroup J dominates, are shown to be related to a written language of ancient Urartu, the political state that was located in the north of the Fertile Crescent (Diakonoff & Starostin, 1986). This is a strong argument that suggests that Caucasians bearing haplogroup J expanded to their current range from the Fertile Crescent. Gavashelishvili & Tarkhnishvili (2016) suggest that this mainly happened in the early Holocene, when natural barriers between the human refugia of Fertile Crescent and West Caucasus disappeared as a result of post-glacial warming. They showed rough correspondence of individual Y-chromosome DNA haplogroups distribution to the refugia, modeled on the basis of paleovegetation data and known human locations from Last Glacial Maximum. Hence, the current pattern of the Y-chromosome DNA haplogroup distributions in the Caucasus may be inherited from the early post-glacial period, perhaps the Paleolithic or early Neolithic. It appears that so far the integration of Caucasian populations into ethnic and political units has not completely erased genetic differences between these populations. This is evident from differences in the frequency of Y-chromosome DNA lineages between local populations of Georgians (Tarkhnishvili et al., 2014), even in spite of considerable gene flow among local populations, especially since industrial era in the XIX-XX centuries. In spite of a clear pattern of Y-chromosome DNA haplogroup differentiation in the Caucasus largely maintained through the combination of male territoriality and landscape heterogeneity, autosomal markers are expected to exhibit a higher degree of admixture due to female-biased gene flow (Sielstad et al., 1998; Nasidze et al., 2004a).

In this study, we tested whether the genetic structure of differentiated human populations could be detected for a small region between the Greater Caucasus and Eastern Turkey, using a limited set of autosomal microsatellite loci. We characterized the factors accounting for the observed patterns, testing ethnicity, linguistic differences, geographic distance, and landscape structure. We also compared autosomal differentiation among the human populations with Y-chromosomal differentiation. For this

purpose, we compared autosomal and Y-chromosome DNA profiles of Georgians from four provinces of the country, as well as Armenians of Eastern Anatolian origin. We showed how the geographic landscape within a limited geographic space has determined the processes of population expansion and limiting gene flow in the past, and has helped to maintain the resulting pattern until the present.

Material and Methods

Sampling. We collected DNA samples from five geographically and linguistically distinct groups from six historical provinces of Georgia (Fig 1): (1) 25 Georgians from the mountain province of Tusheti in the eastern Greater Caucasus, speaking a distinct local dialect of Georgian language (Kartvelian linguistic family) and overwhelmingly represented by Y-chromosome DNA lineage J2 (Tarkhnishvili et al., 2014); (2) 26 Georgians from the mountain province of Svaneti in the western Greater Caucasus, speaking a distinct language of the Kartvelian linguistic family and overwhelmingly represented by Y-chromosome DNA lineage G2a; (3) 32 Georgians from the country's western province of Samegrelo (mostly in lowland or foothills of the Greater Caucasus), speaking a distinct language of the Kartvelian linguistic family and mostly represented by the haplogroups J2, G2a, R1b, L; (4) 30 Georgians from the country's southern province of Meskheti in the Lesser Caucasus, speaking a local dialect of Georgian (Kartvelian linguistic family) and (5) 21 Armenians from the country's southern province of Javakheti, descendants of the families displaced from Mush and Erzurum provinces of eastern Turkey in the early 19th century (Fig 1). The samples were collected from locals with no ancestors from outside of the respective ethnic/geographic population over the last three generations. DNA was extracted from10-12 male chest hairs. Extraction was performed using Qiagen DNeasy Blood and Tissue kit, following the manufacturer's recommendations (QIAGEN, Valencia, California, USA).

Ethics Statement. The research team members, through their contacts in the studied communities, enquired whether locals would voluntarily participate in genetic research that would enable them to retrace their remote ancestry based on their Y-chromosome DNA haplogroup identification. There was a verbal agreement made with volunteer donors of DNA samples, according to which the results would be communicated, electronically or in hard copy with participants individually. Participants were informed that, upon the completion of the lab study, the research would be published without mentioning the names of sample donors. Those who agreed provided us with the envelopes containing their chest hairs, with the birth place of their ancestors (last 3 generations) written on the envelope. In accordance with the preferences of the sample donors, the agreement was verbal and not written. The envelopes are stored as evidence of voluntary provision of the samples and the related information. Analysis of data was done anonymously, using only location and ethnic information; only the first author of the manuscript had access to names associated with samples. Therefore, this study was based on non-invasive and non-intrusive sampling (volunteers provided hair samples they collected themselves), and the information destined for open publication does not contain any personal information. The study methodology was discussed in detail with the members of the Ilia State University Commission for Ethical Issues before the field survey started, and it was decided by the Commission that the formal ethical approval was not needed for conducting this study. This is confirmed in a letter from the Commission chairman, a copy of which has been provided to the journal editor as part of the submission process.

PCR, amplification, and visualization. All samples were genotyped for 16 nuclear STR loci using the PowerPlex® 16 HS System. The PowerPlex® 16 HS kit is used for human identification applications including forensic analysis and relationship testing (Levadokou et al., 2001; Ensenberger et al., 2010). The system allows co-amplification and three color detection of sixteen loci (fifteen STR loci and

Amelogenin), including Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818 (16 HS technical manual Promega (Edwards et al., 1992; Bacher et al., 2000; Ensenberger et al., 2010)). Only two pairs of the selected STR markers are located on the same chromosome: CSF1PO + D5S818 (chromosome 5), and D21S11 + Penta D (chromosome 21), 11 autosomal markers are all located on different chromosomes (Edwards et al., 1992), and the 16th marker is used for sex determination. PCR was run with PowerPlex 16 HS System kit according to the following protocol: 10 µl total volume, with 4–6 µl template DNA, 2 µl PowerPlex® HS 5X Master Mix, 1µl PowerPlex® 16 HS 10X Primer Pair Mix and Amplification Grade water. Thermal cycling was performed at 96 °C for 2 min, 9 cycles of 94 °C for 45 s, 60 °C for 35s, 70 °C for 45 s, 17 cycles of 90 °C for 30 s, 60 °C for 30s, 70 °C for 45 s followed by final 60 °C 10 min. Amplicons were run on an ABI 3130 Automated Genetic Analyzer. Genotypes were screened using GENEMAPPER v. 3.5 (Perkin- Elmer, Waltham, MA, USA).

For identification of Y-chromosome haplogroups, the same samples were genotyped for 23 loci, including DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and Y-GATA-H4. STR markers were amplified according to the following protocol: in 10µl total volume, with 4–6 µl template DNA, PowerPlex. Y235X Master Mix 2µl, PowerPlex. Y2310X Primer Pair Mix 1 µl. Thermal cycling was performed at 95 °C for 2 min, 30 cycles of 94 °C for 20 s, 61 °C for 1 min, 72 °C for 45 s, followed by final extension at 60 °C for 20 min. Amplicons of both autosomal and Y-chromosome markers were run on an ABI 3130 Automated Genetic Analyzer and CC5 Internal Lane Standard 500 Y23. Genotypes were screened using Genemapper v. 3.5 software package (Perkin- Elmer, Waltham, MA, USA). The updated recommendations of the DNA Commission of the International Society of Forensic Genetics for analysis of Y-STR systems were followed (Gusmão et al., 2006). The number of genotypes and haplotypes identified for each geographic population was: 24 of Tusheti, 26 of Svaneti, 20 of

Samegrelo, 30 of Meskheti, 21 Armenians for autosomal loci; 25 of Tusheti, 19 of Svaneti, 32 of Samegrelo, 25 of Meskheti, 21 Armenians for Y-chromosome DNA loci.

Identification of Y-chromosome DNA haplogroups. Y-STR haplotypes were grouped into Y-chromosome DNA haplogroupsE1b1b, I2, J2, J1, G2a, L, T, R1b, using (1) the Y-DNA calculator (Athey, 2005, 2006) and (2) the haplogroup predictor (y-predictor) by Vadim Urasin (available at http://predictor.ydna.ru/). We used the approach described in a previous paper (Tarkhnishvili et al., 2014). The approach results in misclassification rates below 1%. The most common haplogroups were G2a, J2, J1, and R1b (see also Tarkhnishvili et al., 2014).

Analysis of the population genetic structure. We used two different approaches for exploring the structure of the studied populations. The first one was a traditional approach of estimating genetic differentiation between the five studied populations, and in the second approach individuals were assigned to one of the panmictic units, inferred using Bayesian modeling. We estimated genetic differentiation among the populations using Rst values, an analogue of ordinary F statistics commonly applied for microsatellite genotypes, which takes into account the extent of differences between individual alleles at the same locus (Slatkin, 1995; Excoffier, 2001). Rst values were calculated using the software ARLEQUIN v3.5 (Excoffier, Smouse and Quattro, 1992; Excoffier, Laval, and Schneider, 2005). We calculated both Rst based on the autosomal loci and *patrilineal* Rst values (hereafter RstP) using the Y-chromosome DNA STR profiles.

We also measured standard indices of genetic diversity, including observed heterozygosity (autosolmal loci only), mean allele number and range per locus, and Garza-Williamson index (GW) (both autosomal and Y-chromosome DNA loci) using ARLEQUIN v3.5. GW index derives from the ratio of microsatellite allele number and the allele range, where low values of the index indicate a possible

bottleneck that a population has experienced - i.e. it can be used as an indirect measure of the effective population size across time (Garza & Williamson, 2001).

For inferring spatial structure for the five studied samples, we used the matrix of pair-wise Rst and RstP values and their significance taken separately for the autosomal and Y-chromosome DNA markers. We constructed the UPGMA trees based on the genetic differentiation among the five populations, using the free software R (Dray and Dufour, 2007). We tested the statistical significance of differentiation among the tree clusters by calculating Rst/ RstP and P-values among them, using locus-by-locus AMOVA (autosomal loci only).

To separate the entire dataset of genotypes into the groups with a high genetic cohesiveness based on the autosomal markers, software STRUCTURE v2.2 (Pritchard, Stephens, and Donnelly, 2000) was used. Markov chain Monte Carlo parameters were set to a burn-in period of 10,000, and 100,000 replications after burn-in. We repeated the procedure ten times for each set of *a priori* delimited number of clusters (K) ranging from 1 to 6, in order to calculate a probability for each K as described in Pritchard, Wen, and Falush (2009), with the default parameter set (no prior information included).

Landscape heterogeneity & genetic structuring. We used the isolation-by-resistance (IBR) model to incorporate data on landscape heterogeneity into genetic structuring of our study populations. IBR is based on the resistance distance that, as a predictor of genetic differentiation, is likely to perform better than Euclidean or least cost path-based distance measures (McRae, 2006). Unlike Euclidean and least-cost distances, the IBR algorithm assumes that a disperser does not have complete knowledge of the landscape it is traversing and can use multiple paths to reach a destination. The resistance distance was calculated from relationships between random walk times and effective resistances in electronic networks via the software Circuitscape 4 (McRae, Shah & Mohapatra, 2013). The program models multiple random walk paths between populations across a resistance grid that is a raster map, wherein

the value of each cell indicates the relative difficulty (or resistance) of moving through that cell. The program treats the raster map as an electrical circuit, where cells with finite resistances are converted to nodes, cells with infinite resistance (absolute barriers) are dropped, and adjacent cells with zero resistance are consolidated into a single node. In this electrical circuit, adjacent nodes are connected by resistors, with resistances equal to the mean of cell values between a pair of orthogonal neighbors, and the mean resistance multiplied by the square root of 2 between a pair of diagonal neighbors to reflect the greater distance between cell centers. By injecting 1 ampere (by default) of current into each focal node (population) and using Kirchoff's and Ohm's laws, the program calculates effective resistances (i.e. resistance distances), current, and voltages that can then be related to ecological processes (e.g. individual movement and gene flow). Effective resistance acts as the ecological concept of effective geographic distance between populations, but incorporates multiple paths and scales linearly with equilibrium genetic differentiation. The current through nodes or resistors can be ecologically interpreted as expected net movement probabilities (i.e. movement rates) for random walkers moving through a heterogeneous landscape. Voltages can be used to predict the probability that random walkers leaving any point on a heterogeneous landscape will reach a given destination before another.

To see whether landscape complexity acted as a barrier to gene flow among the populations, we performed a correlation analysis between a matrix of genetic differentiation (Rst and RstP) and those of resistance distances (Table S3) among major cultural and economic centers of these populations (Fig 1). The effect of interactions among landscape variables on gene flow was tested using the products of these variables (Table S3). Significance of these relationships was estimated through Mantel Tests at 10,000 permutations, using the free software R (Dray and Dufour, 2007) and its package ade4(R Core Team, 2015). Also, we inferred and mapped movement rates among the five human populations, using the software Circuitscape 4.

Results

Genetic diversity and differentiation: autosomal loci. Genetic profiles of the study populations, both autosomal and Y-chromosomal, are presented in Table S1. Pairwise genetic differentiation between the studied populations (Rst) based on the autosomal STR loci are shown in Table 1a. The highest genetic differentiation was observed between the Tusheti and the Svaneti populations (both ethnically Georgian). All other populations, including the Armenians from Mush and Erzurum regions of Turkey, did not show significant differentiation from each other, although some of them showed significant differentiation from the Tusheti and/or the Svaneti populations.

The Svaneti, the Tusheti, and the Meskheti populations had lower observed heterozygosity than the two other populations. The Garza-Williamson index did not significantly differ among the populations (Table 2).

Genetic diversity and differentiation: Y chromosome. The composition of the haplogroups was significantly different between the studied populations. In the Svaneti population, haplogroup G2a strongly dominated, whereas in the Tusheti population haplogroup J2 was the dominant one. In the rest of the populations, more than one haplogroup with high frequencies were present, with the most common being G2a, J2, R1b, J1, and L, but in different proportions (J2 and G2a dominating in the Samegrelo population; J2 and R1b in the Armenian population; J1, J2 and R1b in the Meskheti population) (Fig 1). Similar to the data based on the autosomal loci, the highest patrilineal difference was between the Tusheti and Svaneti populations, as well as between these two populations and most of the others (Table 1b).

Patrilineal genetic differentiation between the populations (RstP), are shown in Table 1b. The Svaneti population had the lowest number of alleles per locus, and the lowest allele range. The

Samegrelo population had the highest number of alleles per locus and the highest Garza-Williamson index. The Tusheti population had the lowest Garza-Williamson index (Table 2b).

The correlation between pair-wise autosomal and patrilineal differentiation was minor and insignificant (Mantel-Test, r =0.208, p >> 0.05).

Spatial pattern. A UPGMA tree based on the autosomal loci showed a high degree of separation between the Tusheti-Samegrelo cluster and the cluster including the other three populations (Fig 2a). The Rst value was significantly different from zero when these two clusters were compared, but was insignificant when individual populations or population groups were compared within each of the two clusters.

The UPGMA tree, based on patrilineal differentiation, showed a different configuration from that based on autosomal loci (Fig 2b). Specifically, the tree showed high patrilineal similarity between the Svaneti and the Samegrelo populations, which strongly differ if autosomal loci are considered, and the Tusheti population clustered with closely related populations from Meskheti and Armenians. The RstP values were highly significant when all population pairs and clusters were compared to each other, except the Samegrelo and the Svaneti populations.

The Bayesian modeling with STRUCTURE, without including prior information on the sampling locations showed the highest support for K=2 (TableS2). The analysis showed some differentiation among the studied populations (Fig 3). The average proportion of the conventional cluster 1 was the lowest (0.348+/-0.035) for the Tusheti and the highest (0.6343+/-0.040) for the Svaneti population. The Meskheti population (0.426+/-0.052), the Armenians population (0.553+/-0.051) and the Samegrelo population (0.522+/-0.053) held intermediate positions. One-way ANOVA showed significant ($p < 0.0001$) differences among the groups relative to cluster 1, with pairwise significant differences (Bonferroni test)

between the Tusheti and the Svaneti populations (p =0.0001), the Tusheti population and the Armenians (p =0.037), and the Meskheti and the Svaneti populations (p =0.011).

Genetic vs geographic differences. The values of landscape resistance distances among individual populations are shown in Table 1d. Mantel Test showed significant ($p < 0.05$) correlation of autosomal genetic differentiation (Rst) with the resistance values calculated using the combination of ruggedness, snow cover, and tree canopy cover. Fig 5 highlights movement rate and corridors among the five human populations. For paternal genetic differentiation (RstP), the highest significant correlation was with the resistance values calculated from snow cover time series (Table 4, Fig 4). Neither patrilineal nor autosomal genetic differentiation significantly correlated with simple Euclidean distances among the populations.

Discussion

The output of the analyses of both Y-chromosome DNA and autosomal markers presented here converged on the same conclusion, that landscape permeability to humans, rather than simple geographic distance, is critically important for maintaining the genetic differences among the populations across millennia in the study area. The permeability of the landscape to human populations is accounted for by the interaction of ruggedness, snow cover and tree canopy cover. Snow cover was the most important landscape feature impeding paternal movements. As a result of harsh environmental conditions and challenging terrain providing a stronghold effect, populations from the Greater Caucasus Mountains have maintained genetic signatures that formed in pre- or early post-glacial time, whereas the populations from areas south of the Greater Caucasus have undergone substantial admixture in the following millennia.

Our analyses demonstrated that both autosomal and Y-chromosomal genetic differentiation was the highest (Rst=0.108, RstP=0.348) between the Tusheti and the Svaneti populations, one from the east, and the other from the west of the Greater Caucasus, respectively, and the genetic differentiation among the populations from areas south of this mountain system were relatively low. STRUCTURE simulations based on the autosomal markers, without considering population information as a prior, also showed the highest differentiation between the Tusheti and the Svaneti populations: the conventional cluster 1 had 63% of the Svaneti population gene pool and only 35% of the Tusheti population gene pool (with intermediate values for the other three populations), and the differences were significant. The proportions of the Y-chromosome DNA haplogroups are also sharply different: the majority of the Svaneti population has the haplogroup G2a, whereas the majority of the Tusheti population has the haplogroup J2, and the haplogroup G2a is effectively absent from this population; other populations have both G2a and J2 haplogroups in different proportions. The genetic differentiation among the populations, both based on the autosomal and patrilineal genes, significantly correlated with landscape resistance distances among their geographic centers, determined by ruggedness of terrain, snow cover, and closed-canopy forests. In other words, landscape permeability to migrations has accounted for gene flow rates between the populations rather than ethno-linguistic barriers.

As already mentioned in the introduction, the contrast between the compositions of major Y-chromosome DNA haplogroups for the Western and the Eastern Greater Caucasus was described earlier for both the northern (Marchiani et al., 2008; Balanovsky et al., 2011; Yunusbayev et al., 2012) and the southern (Tarkhnishvili et al., 2014) sides of this mountain range. This paper suggests that a similar transition also exists for the autosomal genes. We relate this pattern with human expansion influenced by landscape complexity in early post-glacial times from the Fertile Crescent area to the Eastern Caucasus (while the Western Caucasus was largely populated by people that settled there before or

during glacial period (Tarkhnishvili et al., 2014; Gavashelishvili and Tarkhnishvili, 2016), and limited gene flow among the Greater Caucasus settlements.

In the introduction of this paper, we reviewed the present knowledge and hypotheses on the formation of human genetic patterns in the Caucasus. Our results suggest that, while gene flow remained low within the Greater Caucasus Mountains, and caused a distinct geographic pattern in this region, it was high in the post-Neolithic period among most of the West Asian populations. A recent study by Busby et al. (2015) also demonstrated a high degree in admixture throughout West Eurasia, but without investigating the situation in remote mountain regions. For this reason, the genetic differentiation among the populations from the areas south of the Greater Caucasus is less distinct.

An earlier study (Yunusbayev et al., 2012) did not identify significant autosomal genetic differentiation between populations of the eastern and the western Greater Caucasus, even though it included 214 samples with the Illumina 610 K SNP markers, along with the 906 samples from throughout the world (Li et al., 2008; Behar et al., 2010. This study also, described relatively little differentiation in autosomal markers within the Caucasus region. The authors did not discuss the genetic cline going along the Greater Caucasus range east to west. Our study suggests the presence of concordant clines, both in autosomal genotypes and Y-chromosome haplotypes, going from the east to the west of the Greater Caucasus range.

Previous studies suggest that the differentiation between human populations may indeed be strongly associated with the landscape permeability, showing higher genetic differentiation of mountain populations compared to lowland ones. In one of the rare cases where differences were found between neighboring lowland populations, Palo et al. (2009) showed significant autosomal differences (Fst, 10 STR loci) between three parts of Finland, one of which went through a narrow bottleneck in late medieval time. However, the differences were very low (Fst ~ 0.003), much lower than the differences described here, and resolution was achieved by using very large sample numbers (over 1000). The study

based on the autosomal 15 STR loci in three groups of Tunisian Berbers (Hodjet-El-Khil et al., 2008) showed significant differentiation between one of the studied populations and the two others, and the pair-wise Fst values varied between 0.02-0.04. Significant differences among mountain populations are more common. Significant autosomal differences exist among the target Basque populations and between the Basques and other Europeans, but not among the rest of European ethnic groups (Zlojutro et al., 2006). Even more prominent differences were recorded among populations of the Himalayas, whose steep and snowy slopes can potentially be no less effective barriers to human admixture than the mountains of the Greater Caucasus. The analysis of linguistically distant populations of the Himalayas, based on the 21 autosomal STR loci (Kraaijenbrink et al., 2014), showed significant differences ($P<0.0001$) between Tibeto-Burman and Indo-European speaking populations, both from southwestern slopes of the Himalayas, but separated by the branches of the main mountain backbone. In contrast, Dravidian and Indo-European speakers, whose populations are located south of the Himalayan Range in relatively gentle and clement terrain, did not show significant autosomal differences, although they speak distinct languages, have different dominant paternal lineages (Kayser et al., 2001), and geographic distances between the groups are much larger. To generalize the pattern, limited accessibility and permeability of mountain regions is an important factor for the persistence of relic genetic structures of humans.

A plausible scenario of formation of the human genetic pattern in the rural areas of the Caucasus is described in Tarkhnishvili et al. (2014) and Gavashelishvili and Tarkhnishvili (2016). The recent publication describing post-glacial ancient (Paleolithic) DNA from western Georgia confirms the hypothesis that the migration from the Fertile Crescent to the Caucasus had already started before the expansion of agriculture from the Middle East: the human remains had haplogroup J (Jones et al., 2015) originated from the Fertile Crescent region. An earlier paper by Tofanelli et al. (2009) also suggests the formation of the genetic structure throughout West Asia in the Pleistocene and Early Holocene,

challenging widespread hypotheses that assume major expansions in historical times as critically important for the makeup of regional human population structure. This present paper suggests that the post-glacial expansion from the West Caucasus Refugium (i.e. Colchis) to the Fertile Crescent Refugium was almost as strong as the expansion from the Fertile Crescent to the Eastern Caucasus. This phenomenon is reflected in some genetic affinity of the Svaneti, the Meskheti, and the East Anatolian-Armenian populations with a north-south decreasing cline for haplogroup G2a frequencies. However, after these early post-glacial expansions, the Greater Caucasus Mountains have largely preserved the genetic differentiation that formed 12-15 thousand years ago.

Acknowledgements

The study was financed by an internal grant of Ilia State University. The authors acknowledge Malkhaz Katsitadze for laboratory assistance. Ripan S Malhi and an anonymous reviewer provided important comments on the first draft of the manuscript considered during the revision. Cort L. Anderson corrected the English of the manuscript.

Received 26 April 2016; revision uploaded 6 September 2016.

References

Athey, T. W. 2005. Haplogroup prediction from Y-STR values using an allele-frequency approach. *J. Genet. Genealogy* 1:1–7.

Athey, T. W. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J. Genet. Genealogy* 2:34 –39.

Bacher, J. W., C. Helms, H. Donis-Keller et al. 2000. Chromosome Localization of CODIS Loci and New Pentanucleotide Repeat Loci. *Prog. Forensic Genet.* 8:33–36.

Balanovsky, O., K. Dibirova, A. Dybo et al. 2011. Parallel Evolution of Genes and Languages in the Caucasus Region. *Mol. Biol. Evol.* 28: 2905–2920.

Battaglia, V., S. Fornarino, N. Al-Zahery et al. 2008. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. *Eur. J. Hum. Genet*. 17:820–830.

Behar, D. M., B. Yunusbayev, M. Metspalu et al. 2010. The genome-wide structure of the Jewish people. *Nature* 466:238–242.

Bulayeva, K., L. B. Jorde, C. Ostler et al. 2003. Genetics and population history of Caucasus populations. *Hum. Biol.* 75:837-853.

Busby, G. B., G. Hellentahl, S. Montinaro et al. 2015. The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Curr. Biol*. 25:2518-2526.

Cinnioğlu, C., R. King, T. Kivisild, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–148.

Diakonoff, I. M., and S. A. Starostin. 1986. Hurro-Urartian as an Eastern Caucasian Language. *Münchener Studien zur Sprachwissenschaft*. 12. Munich.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22:1-20.

Edwards, A., H. A. Hammond, L. Jin et al. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–53.

Ensenberger, M. G., J. Thompson, B. Hill et al. 2010. Developmental validation of the PowerPlex® 16 HS System: an improved 16-locus fluorescent STR multiplex. *Forensic Sci. Int.- Gen.* 4:257-264.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 2005:47–50.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.

Excoffier, L. 2001. Analysis of population subdivision, pp. 271–324. In *Handbook of Statistical Genetics*, edited by D. J. Balding, M. Bishop and C. Cannings. John Wiley & Sons, Chichester, UK; 2001.

Garza, J. C., and E. G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* 10:305-318.

Gavashelishvili, A., and D. Tarkhnishvili. 2016. Biomes and human distribution during the last Ice Age. *Global Ecol. Biogeogr.* 25:563-574.

Grugni, V., V. Battaglia, B. H. Kashani et al. 2012. Ancient Migratory Events in the Middle East, New Clues from the Y-Chromosome Variation of Modern Iranians. *PLoS One* 7: e41252.

Gusmão, L., J. M. Butler, A. Carracedo et al. 2006. DNA Commission of the International Society of Forensic Genetics. Commission of the International Society of Forensic Genetics ISFG an update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Legal Med*. 120:191-200.

Hodjet-El-Khil, H., K. Fadhlaoui-Zid, L. Gusmão et al. 2008. Substructure of a Tunisian Berber Population as Inferred from 15 Autosomal Short Tandem Repeat Loci. *Hum. Biol*. 80:435-448.

Hubisz, M. J., D. Falush, M. Stephens et al. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9:1322–1332.

Jarvis, A., H. I. Reuter, A. Nelson et al. 2008. *Hole filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).* Downloaded 12 December 2015.

Jones, E. R., G. Gonzalez-Fortes, S. Connell et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* 6:8912.

Karmin, M., L. Saag, M. Vicente et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25:459-466.

Kayser, M., M. Krawczak, L. Excoffier et al. 2001. An Extensive Analysis of Y-Chromosomal Microsatellite Haplotypes in Globally Dispersed Human Populations. *Am. J. Hum. Genet.* 68:990-1018.

Kraaijenbrink, T., K. J. van der Gaag, S. B. Zuniga et al. A Linguistically Informed Autosomal STR Survey of Human Populations Residing in the Greater Himalayan Region. *PLoS One* 9: e91534.

Levadokou, E. N., D. A. Freeman, M. J. Budzynski et al. 2001. Allele frequencies for fourteen STR loci of the PowerPlex® 1.1 and 2.1 multiplex systems and Penta D locus in Caucasians, African-Americans, Hispanics, and other populations of the United States of America and Brazil. *J. Forensic Sci*. 46:736–61.

Li, J. Z., D. M. Absher, H. Tang et al. 20087. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* 319:1100-1104.

Marchiani, E. E., W. S. Watkins, K. Bulayeva et al. 2008. Culture creates genetic structure in the Caucasus. Autosomal, mitochondrial, and Y-chromosomal variation in Daghestan. *BMC Genet.* 9:47.

McRae, B. H., V. B. Shah, T. K. and Mohapatra. 2013. Circuitscape 4 User Guide. The Nature Conservancy. http://www.circuitscape.org.

McRae, B. H. 2006. Isolation by Resistance. *Evolution* 60:1551-1561.

Nasidze, I., T. Sarkisian, A. Kerimov et al. 2003. Testing hypotheses of language replacement in the Caucasus. evidence from the Y-chromosome. *Hum. Genet*. 112:255–261.

Nasidze, I., D. Quinque, I. Dupanloup et al. 2004a. Genetic evidence concerning the origins of South and North Ossetians. *Ann. Hum. Genet.* 68:588–599.

Nasidze, I., E. Y. Ling, D. Quinque et al. 2004b. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann. Hum. Genet.* 68:205-221.

Palo, J., I. Ulmanen, M. Lukka et al. 2009. Genetic markers and population history: Finland revisited. *Eur. J. Hum. Genet*. 17:1336–1346.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Pritchard, J. K., X. Wen, and D. Falush. 2009. *Documentation for structure software: Version 2.3. http://pritch.bsd.uchicago.edu/structure.html*. Downloaded at 19 December 2015.

R Core Team. 2015. R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/*.2015.

Scozzari, R., F. Criciani, A. Pangrazio et al. 2001. Human Y-chromosome variation in the Western Mediterranean area: implications for the peopling of the region. *Hum. Immunol.* 62:871-884.

Seielstad, M. T., E. Minch, and L. L. Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* 20:278–280.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Tarkhnishvili, D., A. Gavashelishvili, M. Murtskhvaladze et al. 2014. Human paternal lineages, languages, and environment in the Caucasus. *Hum. Biol.* 86:113-130.

Tofanelli, S., G. Ferri, K. Bulayeva et al. 2009. J1-M267 Y lineage marks climate-driven pre-historical human displacements. *Eur. J. Hum. Genet.* 17:1520–1524.

Wei, W., Q. Ayub, Y. Chen et al. 2012. A calibrated human Y-chromosomal phylogeny based on re-sequencing. *Genome Res.* 18:830-838.

Yunusbayev, B., M. Metspalu, M. Järve et al. 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol*. 29:359–365.

Zlojutro, M., R. Roy, J. Palikij et al. 2006. Autosomal STR Variation in a Basque Population: Vizcaya Province. *Hum. Biol*. 78:599-618.

Table 1. Genetic and resistance differentiation among the studied populations: (a) autosomal pairwise genetic differentiation (Rst) with significance values (exact test) shown in parenthesis; (b) Paternal genetic differences (RstP) with respective p-values shown in parenthesis. Significant (P<0.05) differences are in boldface; (c) Euclidean distances; (d) Resistance distances based on the product of ruggedness, snow cover and tree canopy cover, which has the strongest correlation with Rst (see Table 4); (e) Resistance distances based on snow cover time series, which has the strongest correlation with RstP (see Table 4).

(a) Rst	Tusheti	Meskheti	Armenians	Samegrelo
Meskheti	**0.074 (0.000)**			
Armenians	**0.100 (0.000)**	-0.034 (0.991)		
Samegrelo	0.018 (0.108)	0.014 (0.099)	0.022 (0.081)	
Svaneti	***0.108 (0.000)***	0.006 (0.243)	-0.005 (0.477)	**0.036 (0.036)**
(b) RstP	Tusheti	Meskheti	Armenians	Samegrelo
Meskheti	**0.085 (0.000)**			
Armenians	**0.130 (0.000)**	**0.050 (0.036)**		
Samegrelo	**0.116 (0.000)**	**0.067 (0.000)**	**0.070 (0.000)**	
Svaneti	**0.348 (0.000)**	**0.309 (0.000)**	**0.310 (0.000)**	**0.119 (0.000**)
(c) Eucl_dist (m)	Tusheti	Meskheti	Armenians	Samegrelo
Meskheti	306583.700			
Armenians	555573.000	266347.400		
Samegrelo	419534.300	157923.500	290713.800	
Svaneti	332151.100	160303.200	37969.600	112984.600
(d) resist_slp*snw*cnp	Tusheti	Meskheti	Armenians	Samegrelo
Meskheti	2261.875			
Armenians	1861.464	497.095		
Samegrelo	2071.100	658.797	345.020	
Svaneti	2898.440	1586.300	1211.505	1276.085
(e) resist_ snw	Tusheti	Meskheti	Armenians	Samegrelo
Meskheti	2.724			

Armenians	3.307	1.579		
Samegrelo	2.544	1.115	2.357	
Svaneti	4.801	3.378	4.597	2.182

Table 2. Genetic diversity indices for each of the five studied populations. The calculations are based on the autosomal (a) and Y-chromosome DNA (b) microsatellite loci. NA – mean number of alleles per locus; Obs. H – observed heterozygosity; AR – allelic range averaged for the studied loci; GW – Garza-Williamson statistic. Significant (P<0.05) differences are in boldface.

		NA	Obs. H	AR	GW
(a)					
Tusheti	Mean	**6.867**	**0.74555**	**6.6**	**0.92257**
	s.d.	*2.446*	*0.17185*	*3.225*	*0.09212*
Meskheti	Mean	**7.267**	**0.74741**	**6.867**	**0.944**
	s.d.	*2.12*	*0.09437*	*2.748*	*0.08747*
Armenians	Mean	**6.933**	**0.81503**	**6.6**	**0.93611**
	s.d.	*1.792*	*0.12759*	*2.694*	*0.09757*
Samegrelo	Mean	**6.867**	**0.76155**	**6.333**	**0.94939**
	s.d.	*2.503*	*0.13157*	*2.895*	*0.06841*
Svaneti	Mean	**6.867**	**0.71168**	**6.4**	**0.94389**
	s.d.	*2.416*	*0.16496*	*2.947*	*0.09378*
(b)					
Tusheti	Mean	**4.783**		**6.000**	**0.72543**
	s.d.	*1.565*		*2.780*	*0.17797*
Meskheti	Mean	**5.000**		**5.478**	**0.85988**
	s.d.	*1.279*		*3.287*	*0.2028*
Armenians	Mean	**4.565**		**5.261**	**0.82229**
	s.d.	*1.161*		*2.973*	*0.22861*
Samegrelo	Mean	**5.609**		**5.652**	**0.88845**
	s.d.	*1.469*		*2.534*	*0.16834*
Svaneti	Mean	**3.435**		**3.435**	**0.84447**
	s.d.	*0.788*		*1.95*	*0.20438*

Table 3. The probability distribution for each of the five identified conventional clusters (STRUCTURE run that does not include *a priori* information on the sampling locations) at five studied populations.

population	cluster	Mean	
		Statistic	Std. Error
Tusheti	1	0.1181	0.04177
	2	0.7939	0.06497
	3	0.0032	0.00075
	4	0.0840	0.05759
	5	0.0007	0.00020
Meskheti	1	0.5444	0.07605
	2	0.0124	0.00337
	3	0.0045	0.00089
	4	0.0000	0.00000
	5	0.4387	0.07817
Armenians	1	0.9140	0.01719
	2	0.0750	0.01779
	3	0.0080	0.00133
	4	0.0000	0.00000
	5	0.0029	0.00105
Samegrelo	1	0.6854	0.05572
	2	0.3081	0.05629
	3	0.0053	0.00085
	4	0.0002	0.00020
	5	0.0010	0.00036
Svaneti	1	0.8097	0.06377
	2	0.0122	0.00296
	3	0.1749	0.06295
	4	0.0003	0.00021
	5	0.0026	0.00088

Table 4. Correlation between genetic differentiation and landscape resistance values among human populations (values significant at a p-value of 0.05 are in bold; see Table S3 for abbreviations).

Landscape resistance	**Autosomal differentiation: r (p-value)**	**Y-chromosome differentiation: r (p-value)**
Eucl_dist	0.416 (0.230)	**0.067 (0.412)**
resist_snw	-0.465 (0.892)	**0.882 (0.011)**
resist_NPPH	0.512 (0.058)	**0.664 (0.033)**
resist_slp	0.612 (0.051)	**0.721 (0.051)**
resist_slp*cnp	**0.722 (0.047)**	0.493 (0.231)
resist_slp*snw	0.564 (0.094)	**0.819 (0.017)**
resist_slp*snw*cnp	**0.744 (0.048)**	0.515 (0.225)
resist_slp*snw*elv	0.627 (0.103)	**0.774 (0.017)**
resist_slp*elv*snw*cnp	**0.731 (0.049)**	0.582 (0.158)

Supporting Information

Table S1. Genetic profiles of the studied individuals (Excel file with profile distributions on Y-DNA and autosomal markers). “YDNA” – Y-chromosomal DNA profiles, “Autosomal” – autosomal STR loci profiles. For each individual, Y-chromosomal DNA profile, the respective haplotype, calculated using Athey (2006) haplogroup predictor is shown with the respective probability * 100. If probability value is below 99, additionally haplogroup scored using Vadim Urasin haplogroup predictor, with the respective probability, is shown.

Table S2. Minimum lnP(D) for different K and probabilities of K (P(K)) varying from 1 to 6, calculated as described in Pritchard, Wen &Falush (2009). Least values of lnP(D) for each K, based on the 10 repeats, are selected.

no prior	lnP(D)	P(K)
K=1	**-6031.6**	**0.47048**
K=2	**-6037.1**	**0.445302**
K=3	-6225.7	0.067542
K=4	-6463	0.006295
K=5	-6481.8	0.005216
K=6	-6482.8	0.005164

Table S3. Variables used for calculating effective resistances in gene flow between human populations in the study area.

Abbreviated variable	**Description**
elv	SRTM elevation grid of 90-m cells (Jarvis et al., 2008)
slp	Ruggedness as average slope(^{0}) calculated in 1 × 1kmrectangular focal neighborhoods of a SRTM elevation grid of 90-m cells
cnp	% of tree canopy cover, extracted from 500-m MODIS data 'MOD44B'[1] and averaged over the 2000-2010 year period
NPP	Yearly Net Primary Production (kg_C/m^2/yr) extracted from 1-km MODIS data 'MOD17A3'[1] and averaged over the 2000-2010 year period
NPPH	NPP*[1-(Canopy cover/100)] = the amount of NPP (kg_C/m^2/yr) produced by herbaceous and non-forested biomes – i.e. more accessible production to humans
Eucl_dist	Straight-line (Euclidean) distance
resist_slp	Resistance distance calculated from slp as a resistor
resist_elv	Resistance distance calculated from elv as a resistor
resist_snw	Between population pair-wise resistance distance with snow cover time series as a resistor: $D = \frac{1}{\sum_{i=1}^{n} \frac{1}{R_i}}$ Where i = 1,…,n 8-day intervals, and R_i is the ith resistance distance between two populations, calculated by considering snow cover as a full barrier (i.e. conductance = 0). Snow cover at 8-day intervals is extracted from 500-m MODIS data 'MOD10A2'[1] over the 2000-2010 year period.
resist_cnp	Resistance distance calculated from cnp as a resistor
resist_slp*elv	Resistance distance calculated from the product of slp and elv as a

	resistor
resist_slp*snw	Resistance distance calculated from the product of slp and resist_snw as a resistor
resist_slp*cnp	Resistance distance calculated from the product of slp and cnp as a resistor
resist_elv*snw	Resistance distance calculated from the product of elv and resist_snw as a resistor
resist_elv*cnp	Resistance distance calculated from the product of elv and cnp as a resistor
resist_snw*cnp	Resistance distance calculated from the product of resist_snw and cnp as a resistor
resist_slp*elv*snw	Resistance distance calculated from the product of slp, elv and resist_snw as a resistor
resist_slp*elv*cnp	Resistance distance calculated from the product of slp, elv and cnp as a resistor
resist_slp*snw*cnp	Resistance distance calculated from the product of slp, resist_snw and cnp as a resistor
resist_elv*snw*cnp	Resistance distance calculated from the product of elv, resist_snw and cnp as a resistor
resist_slp*elv*snw*cnp	Resistance distance calculated from the product of slp, elv, resist_snw and cnp as a resistor
resist_NPP	Resistance distance calculated from the reciprocal of NPP as a resistor
resist_NPPH	Resistance distance calculated from the reciprocal of NPPH as a resistor

[1]NASA: http://www.echo.nasa.gov/reverb/about_reverb.htm

Figure captions

Fig. 1. Geographic centers of origin of the populations used in this study (see text for further explanations), and composition of Y-chromosome DNA haplogroups in each of the populations.

Fig. 2. UPGMA trees based on pairwise autosomal (a) and Y-chromosomal (b) differentiation at the STR loci. The numbers at the nodes indicate probability that two populations or clusters belong to the same panmictic population.

Fig. 3. The output of STRUCTURE analysis with the most probable number of clusters evaluated using the procedure of Pritchard et al. (2009). No *prior* information on the sample locations included, K=2. The numbers indicate average individual probability to be associated with conventional cluster 1 for each population.

Fig. 4. Plot of autosomal (Rst) differences among the study populations against resistance distance based on the product of ruggedness, snow cover and tree canopy cover (resist_slp*snw*cnp) (upper panel) and plot of paternal (RstP) differences among the study populations against resistance distance based on snow cover (resist_snw) (see Table 4).

Fig. 5. Isolation-by-resistance (IBR) model of cumulative current flow (analogous to probability of gene flow or migration rate) highlights potential corridors among the study human populations, using the product of ruggedness, snow cover and tree canopy cover as a resistor to human movement. Brighter color predicts higher migration rate. Pie diagrams show population centers, colors indicate haplogroups as in Fig. 1.

Fig. 1

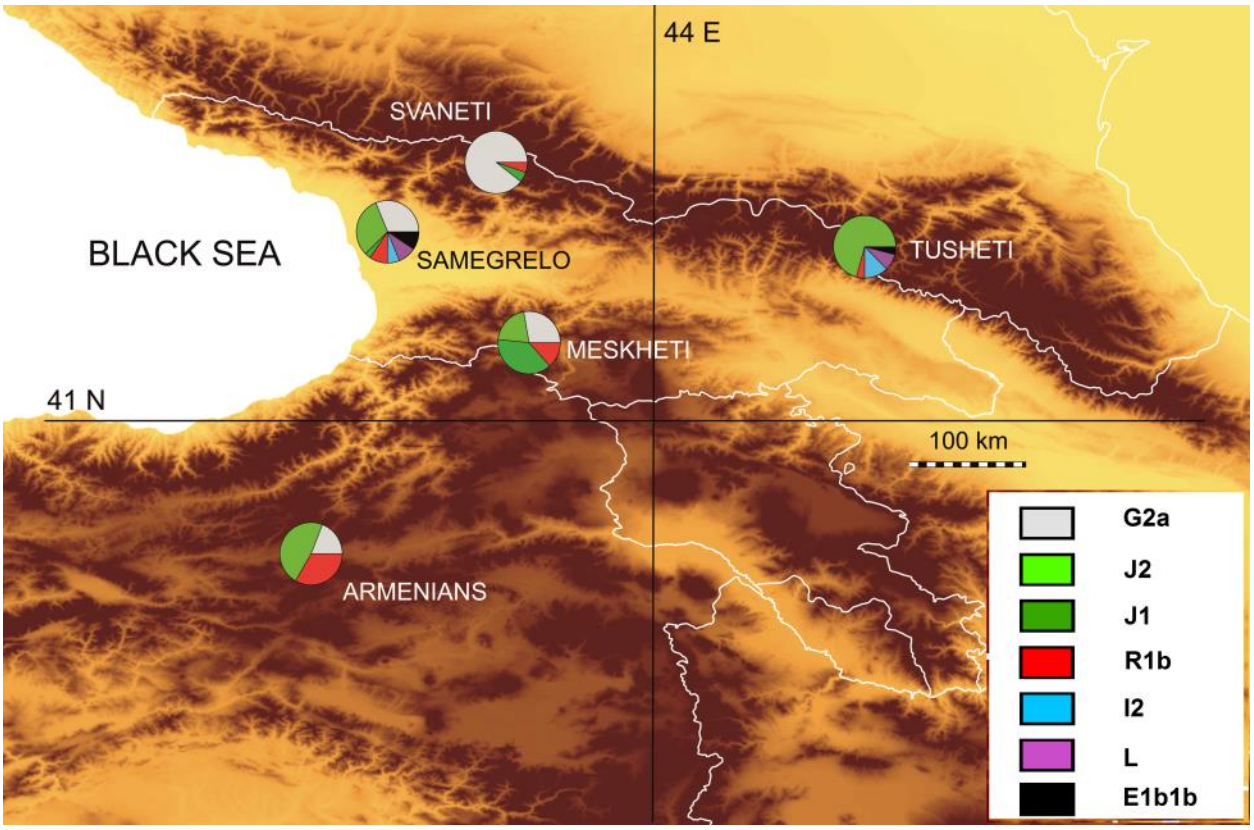


Fig. 2

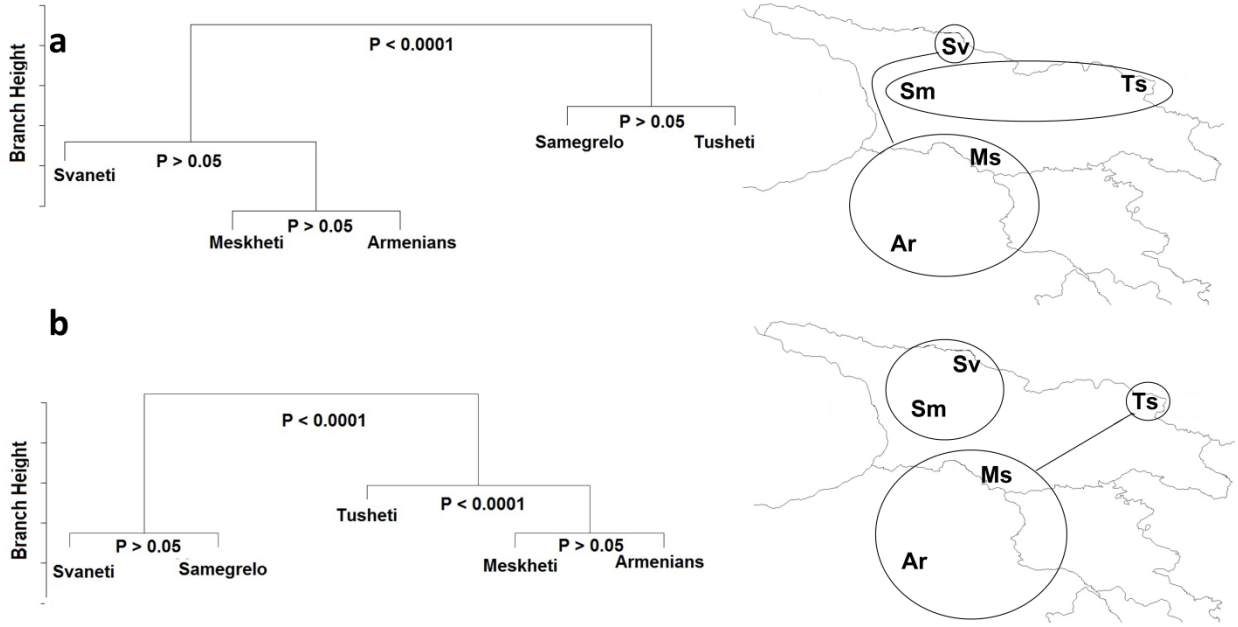
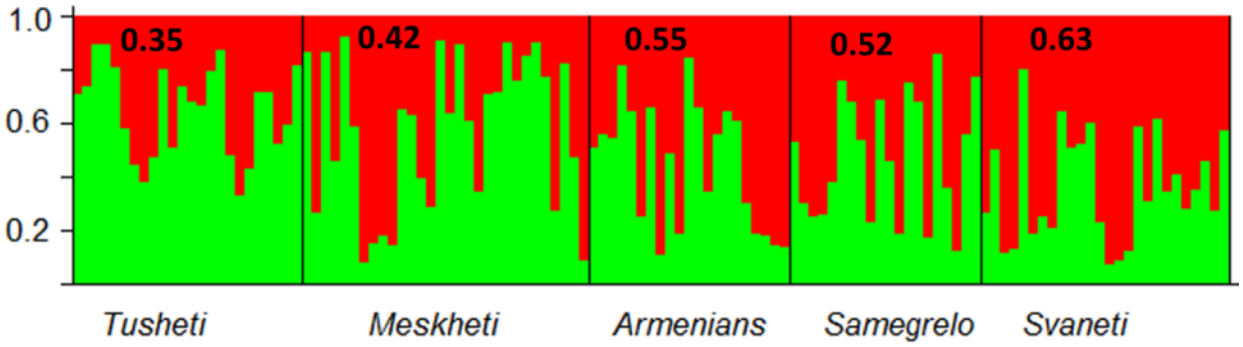
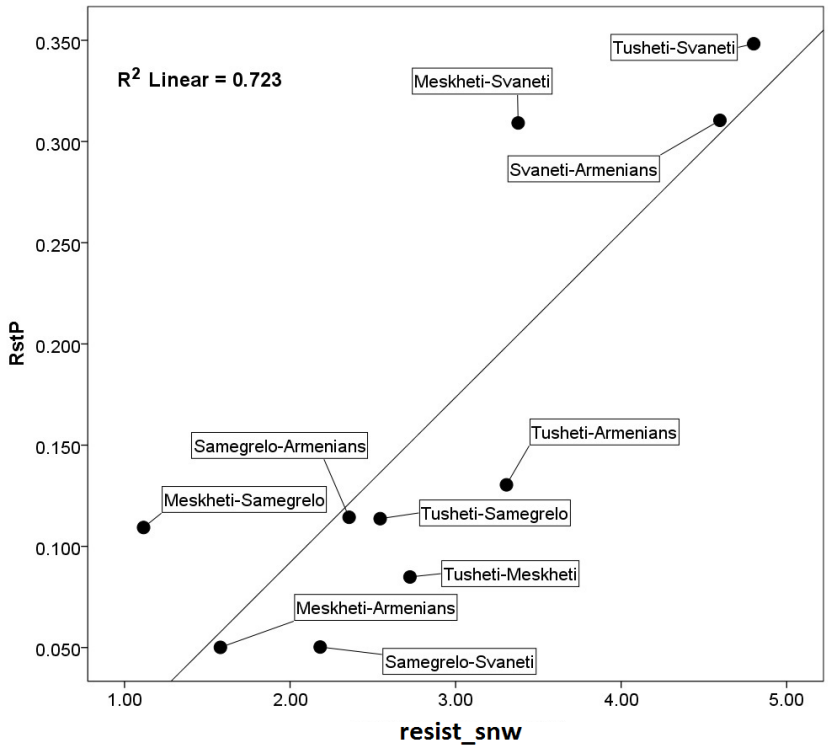


Fig. 3





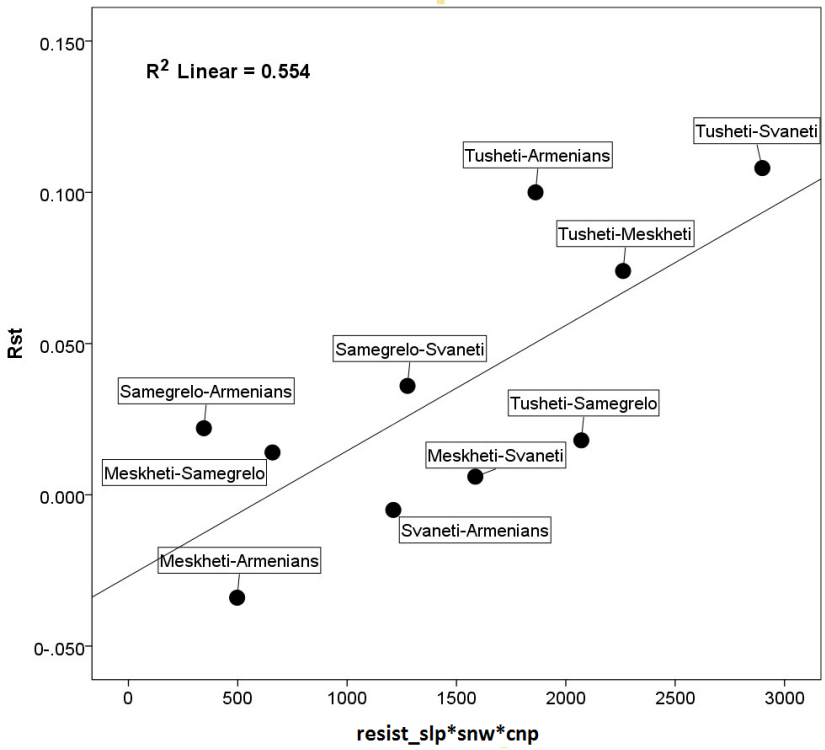


Fig. 4

Fig. 5