Genetic Analysis of Mingrelians Reveals Long-Term Continuity of Populations in Western Georgia (Caucasus)

Theodore G. Schurr [1,*,†], Ramaz Shengelia[2,†], Michel Shamoon-Pour [3], David Chitanava [4], Shorena Laliashvili[4], Irma Laliashvili[4], Redate Kibret[5], Yanu Kume-Kangkolo[6], Irakli Akhvlediani [7], Lia Bitadze[4], Iain Mathieson [8], and Aram Yardumian [5,*]

[1]Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[2]Department of the History of Medicine, Tbilisi State Medical University, Tbilisi, Georgia
[3]First-year Research Immersion, Binghamton University, Binghamton, New York, USA
[4]Laboratory for Anthropologic Studies, Ivane Javakhishvili Institute of History and Ethnology, Tbilisi, Georgia
[5]Department of History and Social Science, Bryn Athyn College, Bryn Athyn, Pennsylvania, USA
[6]Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[7]Georgian and Caucasus YDNA, Family Tree DNA, Houston, Texas, USA
[8]Department of Genetics, Perelman School of Medicine, Philadelphia, Pennsylvania, USA

[†]These authors are co-first authors.

*Corresponding authors: E-mails: tgschurr@sas.upenn.edu; aram2@sas.upenn.edu.

Accepted: October 28, 2023

Abstract

To elucidate the population history of the Caucasus, we conducted a survey of genetic diversity in Samegrelo (Mingrelia), western Georgia. We collected DNA samples and genealogical information from 485 individuals residing in 30 different locations, the vast majority of whom being Mingrelian speaking. From these DNA samples, we generated mitochondrial DNA (mtDNA) control region sequences for all 485 participants (female and male), Y-short tandem repeat haplotypes for the 372 male participants, and analyzed all samples at nearly 590,000 autosomal single nucleotide polymorphisms (SNPs) plus around 33,000 on the sex chromosomes, with 27,000 SNP removed for missingness, using the GenoChip 2.0+ microarray. The resulting data were compared with those from populations from Anatolia, the Caucasus, the Near East, and Europe. Overall, Mingrelians exhibited considerable mtDNA haplogroup diversity, having high frequencies of common West Eurasian haplogroups (H, HV, I, J, K, N1, R1, R2, T, U, and W. X2) and low frequencies of East Eurasian haplogroups (A, C, D, F, and G). From a Y-chromosome standpoint, Mingrelians possessed a variety of haplogroups, including E1b1b, G2a, I2, J1, J2, L, Q, R1a, and R1b. Analysis of autosomal SNP data further revealed that Mingrelians are genetically homogeneous and cluster with other modern-day South Caucasus populations. When compared with ancient DNA samples from Bronze Age archaeological contexts in the broader region, these data indicate that the Mingrelian gene pool began taking its current form at least by this period, probably in conjunction with the formation of a distinct linguistic community.

Key words: Samegrelo, haplogroup, mtDNA, Y-chromosome, autosomal DNA.

Significance

Genomic studies of Caucasus populations have tended to reveal a complex migration history linked to the modern human settlement of western Eurasia, although leaving unanswered questions about the peopling of the Caucasus region itself. Our genetic analysis of Mingrelians from western Georgia showed them to have significant mitochondrial DNA, Y-chromosome, and autosomal diversity, while also revealing subtle differences between North and South Caucasus populations, both modern day and prehistoric. Contextualized with data from Bronze Age individuals, our results suggest that Georgian populations likely emerged as a distinct gene pool in the Bronze Age and were influenced by later expansions of populations into the region.

Introduction

The Caucasus is the seemingly obvious route from two strongly suspected incubators of anatomically modern human diversity (Arabia, Anatolia) to the Eastern European steppe (Ukraine, western and southern Russia) and beyond and has played an important role in the modern human dispersal and settlement in Eurasia since at least the Upper Paleolithic. In this regard, the Caucasus may have been an incubator of human genetic diversity in addition to being a conduit for human migration. Landmark population studies focusing on this region (Barbujani et al. 1994; Nasidze et al. 2004; Balanovsky et al. 2011; Yunusbayev et al. 2012; Yardumian et al. 2017) have revealed great and unusual diversity that points to a complex peopling process for the Caucasus that is related to those of the Near East, the Pontic steppe, Central Asia, and Europe. High mitochondrial haplotypic diversity among even putatively endogamous highland populations such as Svans (Yardumian et al. 2017) suggests multiple human introgressions into the region rather than significant founder effects. The tight genetic clustering between populations speaking languages from different families (e.g., Abkhaz, Svan, and Ossetian) further points to the continuity of lineages from a time preceding linguistic diversification (Yunusbayev et al. 2012; Yardumian et al. 2017). Despite these general observations, it is still not clear to what extent contemporary North and South Caucasus populations descend from those of the Bronze Age, Neolithic, and earlier periods of time.

For millennia, humans have found Samegrelo (Mingrelia or Megrelia) and neighboring regions of Georgia suitable for settlement and transhumance due to its favorable climate and abundant resources. The coastal lowlands of Georgia and the North Caucasus are an easily traversable route from the world of the Southern Arc to the world of the Eastern European steppe, unimpeded by the Caucasus Mountains (Vasilyev and Amirkhanov 2018). Western Georgia also features a dense network of rivers running from the mountain glaciers to the Black Sea. In addition to fresh river water and Black Sea marine resources, minerals, metal ores, and obsidian have long been harvested in the nearby Caucasus highlands (Courcier 2014), while a variety of flora and fauna are found on the fertile Colchian Plain (Grossheim 1952; Agladze et al. 1998).

Linguistically, Mingrelian is one of four extant Kartvelian languages. The Kartvelian family includes Georgian, Mingrelian and Laz, which are closely related, and Svan, which is an older member of the family (Boeder 2005; Chirikba 2008) (supplementary fig. S1, Supplementary Material online). Various theorists have posited the South Caucasus as a point of origin for the emergence of Kartvelian languages (Klimov 1998), although others suggest they were also once spoken in Anatolia (Diakonoff and Starostin 1986; Kassian 2010; Oštir 1921; Kavtaradze 1983, 2002). Some historians and archaeologists (e.g., Javakhishvili 1998) believe that Proto-Kartvelian arose in an area encompassing present-day Georgia. Under this scenario, Proto-Kartvelian was likely present in western Georgia by the early Bronze Age, and the Svan and Georgian-Zan languages, representing the deepest branch of Kartvelian, probably separated not long after that (Javakhishvili 1998) (supplementary fig. S2, Supplementary Material online). The closely related Laz and Mingrelian languages appear to have diverged relatively recently, perhaps because of the westward spread of Georgian in the first millennium CE (Kavtaradze 1983).

Today, Mingrelian is spoken throughout the seven municipalities of Samegrelo (Tsalenjikha, Chkhorotsqu, Abasha, Martvili, Zugdidi, Khobi, and Senaki) and, prior to the armed conflict of 1992–1993, in a greater portion of Abkhazia than today (Vamling and Tchantouria 2005). There is no reliable numerical estimate of contemporary Mingrelian speakers, as all Kartvelian speakers are combined as an ethnolinguistic group in the national census (GeoStat 2016). All Mingrelians and Svans are also at least conversant in Georgian (Vamling and Tchantouria 2005). While it has been written with the Georgian alphabet since the mid-19th century, Mingrelian is not taught in schools and does not have a literary tradition (Hewitt 2017).

Given the historical and linguistic complexity of this region, the relationships between the peoples residing on either side of the Caucasus mountains have been the focus of intensive study, as have those between Caucasus populations and populations occupying the East European steppe, the Middle East, and Central Asia (Schönberg et al. 2011; Haber et al. 2016; Wang et al. 2019). A primary question is the extent to which

neighboring peoples of the Caucasus who do not share a language affiliation are genetically related to each other (Nasidze et al. 2004; Balanovsky et al. 2011). Our previous study of Svan genetic history (Yardumian et al. 2017) demonstrated significant mitochondrial DNA (mtDNA) heterogeneity and more limited diversity in the nonrecombining region of the Y-chromosome (NRY) in this population, as well as the close genetic affinity between Svans and neighboring highland populations (Abkhazians and Ossetians) who do not speak Kartvelian languages. This observation suggested that the pattern of genetic variation for Mingrelians could be the same.

Thus, to elucidate the history of Mingrelians and their relationship with populations from the Caucasus and neighboring regions, we conducted an anthropological genetic study in Samegrelo, an important cultural–historical region in western Georgia (fig. 1). Relatively few genetic data from Mingrelians have been previously published (e.g., Roostalu et al. 2006; Khusnutdinova et al. 2012), and these alone were hardly sufficient to detect patterns of genetic variation in this region of Georgia. Our phylogeographic analyses of mtDNA and Y-chromosome data provide a more comprehensive picture of Mingrelian population history from the haploid level, while the analyses of autosomal single nucleotide polymorphism (SNP) diversity reveal Mingrelians to have genetic affinities with other Caucasus populations, with subtle differences between North and South Caucasus groups being evident. Overall, this analysis contributes new insights into the genetic diversity of Mingrelia and expands our knowledge of the history of the culturally diverse Caucasus region in which it is situated.

Results

mtDNA Diversity in Mingrelians

We analyzed mtDNA diversity in 446 Mingrelian-speaking individuals representing 29 communities from across the province of Samegrelo (Mingrelia). We utilized data generated from control region (CR) sequencing and genotyping of mitogenomes with 5,205 SNPs, the majority of which occurred in the coding region. Comparison of the CR sequences and SNP genotypes produced haplogroup (hg) calls that were essentially congruent with each other (supplementary table S1, Supplementary Material online).

Overall, Mingrelians exhibited considerable mtDNA hg diversity. They possessed high frequencies of West Eurasian hgs (H, HV, I, J, K, N1, R1, R2, T, U, and W),

Fig. 1.—A map of the administrative districts of Georgia. Redrawn by Girmaye Misgna based on a color map available at OnTheWorldMap (https://ontheworldmap.com/georgia/map-of-georgia-1500.jpg).

Table 1
mtDNA hg Frequencies in Mingrelians, Svans, and Georgians

	Mingrelians		Svans		Georgians*	
Hg	***n***	**%**	***n***	**%**	***n***	**%**
A	2	0.44	0	—	0	—
C	3	0.66	7	3.80	2	2.67
D	6	1.32	3	1.63	1	1.33
F1	5	1.10	0	—	1	1.33
H	66	14.54	33	17.93	24	32.00
HV	18	3.96	2	1.09	0	—
I1	16	3.52	3	1.63	3	4.00
I3	1	0.22	0	—	0	—
J	11	2.42	1	0.54	3	4.00
K	42	9.25	29	15.76	4	5.33
N1	0	—	1	0.54	1	1.33
N1a	3	0.66	0	—	0	—
N1b	17	3.74	2	1.09	1	1.33
N9	3	0.66	0	—	0	—
R0a1	0	—	1	0.54	0	—
R1	8	1.76	0	—	0	—
R2	7	1.54	3	1.63	0	—
T	0	—	6	3.26	8	10.67
T1	20	4.41	3	1.63	4	5.33
T2	34	7.49	8	4.35	2	2.67
U	0	—	3	1.63	1	1.33
U1	16	3.52	15	8.15	0	—
U2	9	1.98	11	5.98	2	2.67
U3	30	6.61	4	2.17	3	4.00
U4	52	11.45	4	2.17	3	4.00
U5	15	3.30	5	2.72	5	6.67
U7	4	0.88	3	1.63	1	1.33
V	0	—	0	—	2	2.67
W	21	4.63	24	13.04	2	2.67
X2	39	8.59	12	6.52	2	2.67
X4	6	1.32	1	0.54	0	—
Total	**454**	**99.97**	**184**	**99.97**	**75**	**100.00**

Note.—The Svan data were taken from Yardumian et al. (2017). The asterisk (*) indicates that the Georgian data were combined from Yardumian et al. (2017), Nasidze et al. (2004), and this study and represent persons from different regions of the country. Georgians from this study ($n = 6$) had F1c, T1a (2), T2, U3b, and U5a mtDNAs. For individuals excluded from this table: non-Mingrelian Georgian citizens ($n = 3$) had H1b1, K1c, and U4 mtDNAs; Mingrelian speakers with Slavic admixture ($n = 14$) had H, H1b2, H26a1, H28, H5a1 (2), HV1a, IIa1, K1a3, R1a, T1a, T2, U4a, and U5b mtDNAs; and Mingrelian relatives ($n = 7$) had HV1a, I1b, I3, T1a, T2, U1a, and U3 mtDNAs.

including some of the highest frequencies and diversity of X2 and X4 yet detected. Certain East Eurasian hgs (A, C, D, and F) appeared at low frequencies and represented some 3.59% of Mingrelians' mitochondrial gene pools (table 1). Many of these maternal lineages had previously been observed in neighboring Svans (Yardumian et al. 2017). By comparison, the aggregate population of Georgians (i.e., not including Mingrelians and Svans) showed a relatively similar pattern of hg diversity to those seen in both Mingrelians and Svans, despite it being comprised individuals living in several different areas of the country.

We also examined patterns of matrilineal diversity in the Caucasus and surrounding regions. As shown in figure 2, the distribution of mtDNA hgs was relatively similar for populations from the North and South Caucasus, although possibly affected by the sample sizes for them. Among South Caucasus populations, the Svans differed subtly from Mingrelians and Georgians (i.e., exclusive of Svans and Mingrelians) in terms of their high frequency of hg W mtDNAs and lack of hg HV mtDNAs. By contrast, Armenians have a more divergent lineage distribution in which hgs H and HV comprised nearly 40% of their mtDNAs, and East Eurasian hgs were absent.

Phylogeographic Analysis of Mingrelian mtDNA Haplotypes

To further evaluate the mtDNA diversity in Mingrelians, we conducted phylogenetic analyses of specific maternal lineages present in this population using network analysis. We included in our analysis data from comparative populations to better delineate the phylogeography of these hgs. We focused on several lineages that had the potential to reveal insights into the genetic diversity present of Georgian populations, including the widespread hgs U4 and X2 (and their sub-branches) and the East Eurasian hgs C4a1a, D4g2a, and F1b1.

hg U4

U4 was the most common mtDNA hg in Mingrelians (11.1%). Its haplotypes occurred at a higher frequency and greater diversity than seen in Svans. Out of 54 Mingrelian U4 lineages, 28 represented six different haplotypes within the U4a subclade, and 21 (seven haplotypes) belonged to the U4b1b subclade (supplementary table S1, Supplementary Material online). Based on these data, U4a and U4b1b were the most common mtDNA subclades among Mingrelians.

The phylogenetic analysis of U4b1b haplotypes from Mingrelian and populations described in published sources revealed a Georgian branch of the U4b1b1 subclade (supplementary fig. S3, Supplementary Material online). Defined by the 16086C mutation, this cluster was present in only Georgians and a single Armenian (Derenko et al. 2019). The star-like structure of the Georgian U4b1b1 + 16086C subclade suggested that it arose relatively recently in this region. In addition, the two Mingrelian U4b1b2 haplotypes joined two from Svans (Yardumian et al. 2017) to compose a long branch, suggesting some degree of continuity of this subclade within western Georgia.

Overall, the phylogeography of U4b1b points to its deep roots in Europe or the Pontic steppe. While phylogenetic analysis of contemporary lineages suggested a dispersal from the Near East into Siberia and Altai region no later than 4,000 years before present (ybp) (Malyarchuk et al. 2010; Dulik et al. 2012; Derenko et al. 2014), the majority of ancient and contemporary U4b1b haplotypes are reported from continental Europe (https://www.yfull.com/mtree/U4b1b/).

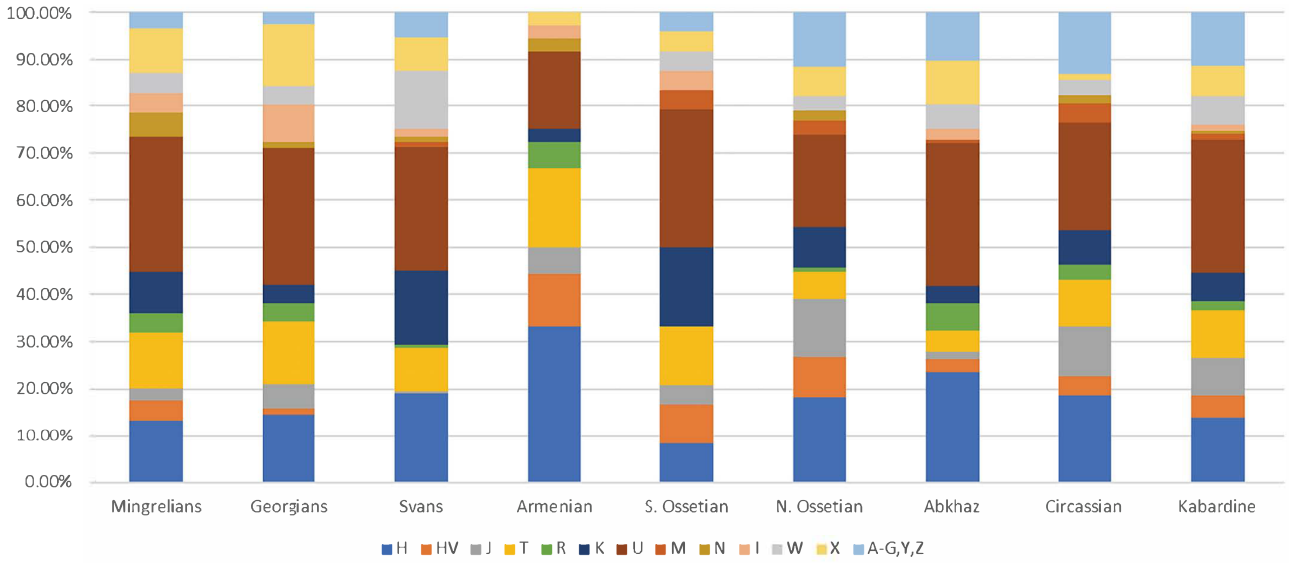

Fig. 2.—mtDNA hg frequencies in Georgians and comparative populations. All East Eurasian hgs (A–G, Y, and Z) have been aggregated into a single hg category because of their relatively limited representation in Mingrelians and neighboring populations. Data sources: Mingrelians = this study ($n = 446$); Georgians = this study, Yardumian et al. (2017), and Battaglia et al. 2009 ($n = 75$); Svans = Yardumian et al. (2017) ($n = 184$); Armenian = Herrera et al. (2012) ($n = 36$); South Ossetian = Yunusbayev et al. 2012 ($n = 24$); North Ossetian = Balanovsky et al. (2011) ($n = 138$); Abkhaz = Balanovsky et al. (2011) ($n = 136$); Circasssian = Balanovsky et al. (2011) ($n = 123$); Kabardine = Balanovsky et al. (2011) ($n = 150$).

Paleogenomic studies have also revealed the presence of U4b1b among Mesolithic and Neolithic individuals from Germany, Serbia, and Ukraine (Allentoft et al 2015; Mathieson et al. 2018; Narashimian et al. 2019; Rivollat et al. 2020; Juras et al. 2021). Based on this evidence, Europe/Pontic region is the likely source of U4b1b haplotypes in Mingrelians.

hg X2

hg X2 mtDNAs occurred in 36 individuals (7.4%) at one of the highest frequencies observed in human populations. The most common subclade among Mingrelians was X2f, which was represented by 17 individuals belonging to ten haplotypes (supplementary table S1, Supplementary Material online). A rare subclade with strong presence in South Caucasus, X2f, has previously been identified in both Svans (Yardumian et al. 2017) and Iranians (Derenko et al. 2013). The rest of the X2 haplotypes appeared in 19 other Mingrelians, the majority of which belonged to three branches of the X2 + 225 clade, primarily subclades X2d and X2e. These subclades have been previously reported in Svans, among whom X2n is also present (Yardumian et al. 2017). As seen for X2f, they are found across the South Caucasus but occur at low frequencies elsewhere in Europe and the Middle East (Reidla et al. 2003; Shlush et al. 2008).

Our network analysis confirmed the great diversity of X2f haplotypes among Mingrelians. Half of the known X2f haplotypes appeared in Mingrelians, with only one being shared with populations from Lebanon and Italy (supplementary fig. S4, Supplementary Material online). With few cases reported from outside the Caucasus, this distribution suggested that X2f originated in this region. In support of this view, the earliest paleogenomic evidence for X2f comes from early Bronze Age Maykop (fourth millennium BCE, Adygea [Russia]) and Kura-Araxes (third millennium BCE, Armenia) sites (Lazaridis et al. 2016; Wang et al. 2019), confirming the presence of this subclade in the Caucasus since at least 5,500 ybp. Given these findings and limited evidence for their distribution into Europe and the Middle East, hg X2f haplotypes appear to have been present in this region at low frequencies since that time.

Supplementary Figure S5, Supplementary Material online shows the phylogenetic relationships between the rest of the Mingrelia hg X2 haplotypes. Nearly all of these haplotypes belonged to subclades X2d and X2e2. As seen with other sub-branches of X2, they have been previously reported in Svans (Yardumian et al. 2017). With 75% of the haplotypes being exclusive to Mingrelians and other Georgians, this network points toward to the continuity of X2d and X2e2 haplotypes in this region of the South Caucasus. Although X2d and X2e2 are rarely seen in West Eurasia, many of their longer branches are represented by individuals from the Caucasus and Iran, suggesting a deeper history of these subclades in these regions. From a paleomitogenomic perspective, X2d haplotypes have been identified

in Neolithic Anatolia (Mathieson et al. 2015) and those for X2d1 in Neolithic Hungary and Germany (Lipson et al. 2017), while the earliest evidence for X2e comes from a sample from Alalakh (ALA04; Hatay Province, Turkey) dated to 1,776–1,853 cal BCE (~3,850 ybp) (Skourtanioti et al. 2020).

X4 is a rare hg with a geographic distribution similar to those of the major Eurasian X2 subclades. It is present at very low frequencies in the South Caucasus and the Middle East, as well as in Europe (particularly the Balkans and Eastern Europe) and Central Asia (Fernandes et al. 2012; Röck et al. 2013). Network analysis of X4 haplotypes revealed considerable diversity in Georgia, with Mingrelian haplotypes being present on all major branches of this lineage and separated from each other by numerous mutations (supplementary fig. S6, Supplementary Material online). The oldest known X4 paleomitogenomes have been identified in individuals from third millennium BCE Scythian (Järve et al. 2019) and Ukrainian Catacomb culture (Juras et al. 2018) sites. Such findings suggest a tentative connection between the Pontic steppe and the South Caucasus.

Aside from these West Eurasia hgs, Mingrelians exhibited a low frequency of mtDNAs belonging to maternal lineages more commonly found in East Eurasia, specifically hgs C4a1a, D4, and F1b1. The networks for these hgs (supplementary figs. S7–S9, Supplementary Material online) revealed Mingrelian haplotypes to occupy positions on the extended branches of these networks rather than the internal nodes. We elaborate on these details below.

hg C4a1a

hg C4a1a appeared in Mingrelians as a single haplotype shared by three individuals. A median-joining network including all available C4a1a sequences from previous studies ($n = 110$) and the Mingrelian individual showed that the vast majority of C4a1a haplotypes appear in Central Asian populations (Kong et al. 2006; Metspalu et al. 2006), although a few were present in Slavic speakers (Mielnik-Sikorska et al. 2013) (supplementary fig. S7, Supplementary Material online). The Mingrelian C4a1a haplotype was closely related to the modal type, which encompasses numerous samples from Central Asia and Eastern Europe and a few from Daghestan and Turkey. Overall, the typology of the C4a1a network suggests a largely westward dispersal of these haplotypes into the Caucasus, possibly through the steppe.

hg D4

While hg D mtDNAs have previously been noted in Svans (Yardumian et al. 2017), the presence of a D4g haplotype in four Mingrelians was curious. The subclade to which these mtDNAs belong, D4g2a, is largely limited to eastern China and specifically Han Chinese and Mongol populations from Russia and northern China (Derenko et al. 2010, 2018; Zhao et al. 2019) (supplementary fig. S8, Supplementary Material online). Ancient DNA evidence also confirmed the presence of this subclade in archaeological samples from eastern China (Henan Province) that dated to over 6 kya (Ning et al. 2020). Interestingly, the D4g2a haplotype in Mingrelian individuals was identical to one reported in a Svan individual (Yardumian et al. 2017) and also two Han Chinese individuals (Loo et al. 2014; Zhang et al. 2014). Although difficult to make phylogeographic inferences about D4g2a in Georgia based on one to two haplotypes, it is nevertheless notable that this subclade has not been reported elsewhere in the Caucasus, Anatolia, or the greater Near East.

hg F1b1

hg F1b1 appeared in six Mingrelians in the form of two haplotypes. F1b1 has previously been reported in China, Nepal, and Vietnam and occurs in the South Caucasus, Iran, and Central Asia at very low frequencies (Schönberg et al. 2011; Derenko et al. 2013). Within the median-joining network of F1b1, the two Mingrelian haplotypes appeared on a branch descending from a haplotype present in both Nepal (Wang et al. 2012) and Yunnan (China) (Zhang et al. 2016) (supplementary fig. S9, Supplementary Material online). Overall, the F1b1 network reflects the diversity of this subclade in Asia, Iran, and the South Caucasus, where it also appears in Armenians, Azeris, and Georgian Jews (Behar et al. 2008; Schönberg et al. 2011). The vast territory in which this relatively rare subclade appears suggests an early dispersal for it. In fact, the earliest paleogenomic evidence of F1b1 comes from two early Neolithic Cis-Baikal individuals whose mtDNAs belonged to F1b1b (Kılınç et al. 2018). Although this analysis does not reveal a clear geographic relationship between the F1b1 haplotypes in Mingrelians and those in other populations, it is likely that they were introduced into Georgia via Iran, given their absence in European Russia and their presence in Central Asia and Iran.

Y-Chromosome Diversity in Mingrelians

We analyzed Y-chromosome diversity in a total of 366 Mingrelian-speaking males. We utilized data generated from Y-short tandem repeat (STR) genotyping using the Yfiler kit and the genotyping of the NRY with 10,272 SNPs. Comparison of the Y-STR haplotypes and SNP genotypes produced hg calls that were congruent with one another (supplementary table S2, Supplementary Material online). As with the mtDNA data, we were able to identify the hg status of most samples to a fine subclade level.

Mingrelians possessed nine major hgs, including E1b1b, G2a, I2, J1, J2, L, Q, R1a, and R1b (table 2). As with Svans,

Abkhazians, and Ossetians (Balanovsky et al. 2011; Yardumian et al. 2017), G2a was the most frequently occurring of these paternal lineages. J2a, which also is found at high frequencies in Iran, Anatolia, and especially among Chechen and Ingush speakers in the North Caucasus, was the next most frequent. Other Western Eurasian hgs, such as E1b1b, J1a, R1a, and R1b, were observed at modest to moderate frequencies. L1b, which was not found in our Svan sample and appeared at very low frequency in certain published Caucasus groups (e.g., Battaglia et al. 2009), occurred at a moderate frequency in Mingrelians. Finally, we detected hgs I2, Q1b, and G1 haplotypes at low frequencies in Mingrelians.

Table 2
Y-chromosome hg Frequencies in Mingrelians, Svans, and Georgians

	Mingrelians		Svans		Georgians	
Hg	***n***	**%**	***n***	**%**	***n***	**%**
E1b1b	12	3.30	0	—	2	3.03
G1	2	0.55	0	—	0	—
G2a	162	44.51	73	78.49	21	31.82
I1	0	—	0	—	1	1.52
I2	3	0.82	4	4.3	0	—
J1	19	5.22	0	—	3	4.54
J2	102	28.02	6	6.45	21	31.82
L1b	25	6.87	0	—	0	—
L3	0	—	0	—	1	1.52
N	0	—	1	1.08	0	—
Q1b	5	1.37	0	—	0	—
R1a	17	4.67	9	9.68	7	10.60
R1b	17	4.67	0	—	6	9.09
R2	0	—	0	—	3	4.54
T	0	—	0	—	1	1.52
Total	**364**	**100.00**	**93**	**100.00**	**66**	**100.00**

Note—The Svan data were taken from Yardumian et al. (2017). The Georgian data were taken from Battaglia et al. (2009). Georgians from this study ($n = 4$) had G2a and J2a (3) Y-chromosomes, while those from Yardumian et al. (2017) ($n = 6$) had E1b1b, G2a (3), J2a, and R1b Y-chromosomes. The two Mingrelian individuals excluded from this table had a G2a (SAM 458) and L1b (SAM 293) Y-chromosomes.

Comparative Y-Chromosome Diversity in the Caucasus and Near East

Mingrelians had a significant frequency of NRY hg G2a, although not as high as previously reported in Svans. Svans had mostly G2a haplotypes, with the rest belonging largely to hgs J2, I2, N, and R1a (Yardumian et al. 2017). Unlike Svans, Mingrelians had moderate frequencies of J2a and L1b Y-chromosomes. Thus, despite some significant overlap in lineage frequencies, Mingrelian and Svan populations had different Y-chromosome profiles.

Overall, Mingrelians showed a pattern of NRY hg diversity similar to those observed in most populations from the North and South Caucasus (fig. 3). Among South Caucasus populations, Svans differed from Mingrelians, Georgians, and other populations in terms of their high

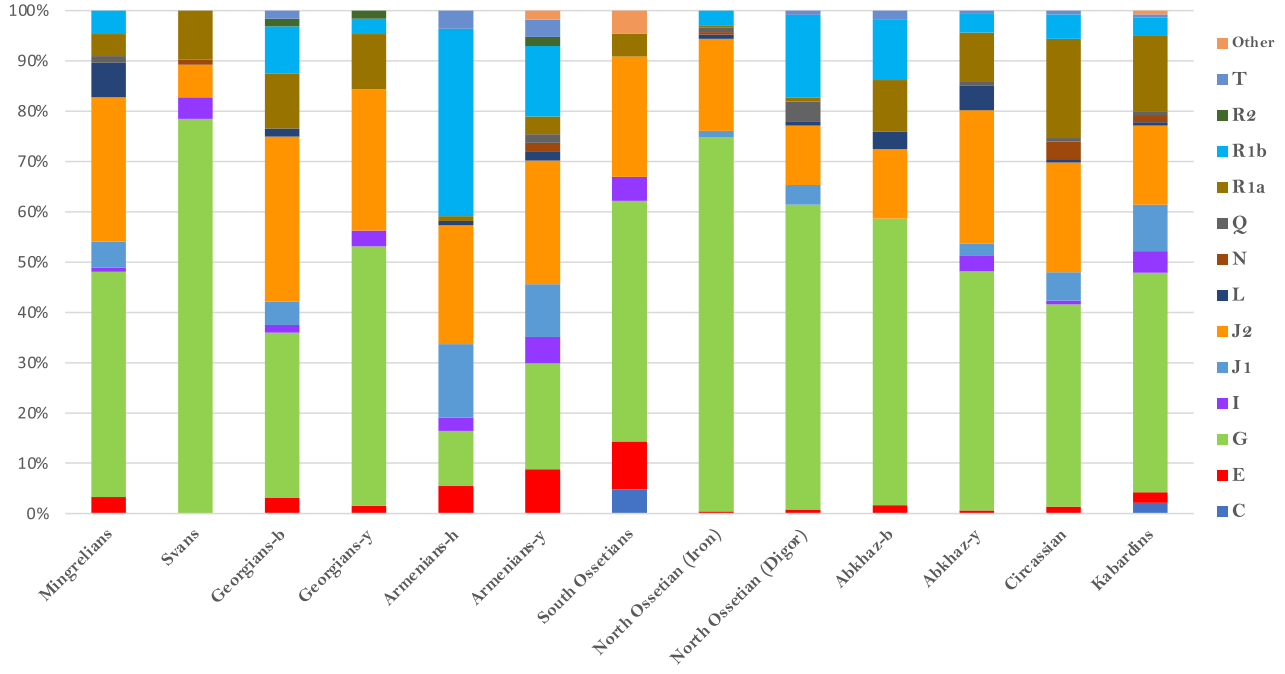

Fig. 3.—NRY hg frequencies in Georgians and comparative populations. Data sources: Mingrelians = this study ($n = 366$); Svans = Yardumian et al. (2017) ($n = 93$); Georgians-b = Battaglia et al. (2009) ($n = 66$); Georgians-y = Yunusbayev et al. 2012 ($n = 65$); Armenians-h = Herrera et al. (2012) ($n = 110$); Armenians-y = Yunusbayev et al. 2012 ($n = 55$); South Ossetians = Yunusbayev et al. 2012 ($n = 21$); North Ossetian (Iron = Balanovsky et al. (2011) ($n = 230$); North Ossetian (Digor) = Balanovsky et al. (2011) ($n = 127$); Abkhaz-b = Balanovsky et al. (2011) ($n = 58$); Abkhaz-y = Yunusbayev et al. 2012 ($n = 162$); Circasssian = Balanovsky et al. (2011) ($n = 142$); Kabardins = Balanovsky et al. (2011) ($n = 140$).

frequency of G2a Y-chromosomes, and Mingrelians had a higher frequency of hg L haplotypes than other groups. North Caucasus populations also had higher frequencies of hg R1a than South Caucasus groups. In addition, Armenians differed from other Caucasus populations in having higher frequencies of hgs E, J1, and R1b, while North and South Ossetians differed somewhat from each other based on their hg C, E, G, and I frequencies.

Phylogeographic Analysis of Mingrelian Y-Chromosome hgs

We initially used Y-STR haplotype data to infer the hg status of Mingrelian samples and determine the phylogeographic status. Once generating SNP data through microarray analysis, we confirmed and elaborated these calls. We further constructed trees/networks with these haplotypes and used information available on the ISOGG website (https://isogg.org) to generate refined phylogenetic trees of lineages present in Mingrelians. Because of their importance for the phylogeographic history of the South Caucasus, we discuss several of these major paternal lineages here.

hg G2a

In the current study of 372 Mingrelian males, 31.45% of them belonged to hg G2a-Z6653 (G2a1a1a-FGC693/Z6653) (supplementary fig. S10, Supplementary Material online). Y-chromosomes belonging to this branch have a time to most recent common ancestor (TMRCA) of 6,521 ybp (https://www.yfull.com/tree/G-Z6653/). The variety of sub-branches within G2a-Z6653 also clearly reflects the deep roots of this hg in the Caucasus, whose estimated age suggests that it arose during the Neolithic period.

Unfortunately, the limited number of archaeological samples from the Caucasus available for ancient DNA analysis has hampered the effort to trace the origin and antiquity of this lineage. In this regard, the oldest ancient samples from the parallel branch for G2a-Z6653 were found in a Neolithic individual from western Iran (dated to 5,837–5,659 BCE) (Lazaridis et al. 2016) and in a Neolithic individual from northern Iraq (dated to 8,300–7,900 BCE) (Lazaridis et al. 2022). These findings suggested a geographic nexus between Iran, Iraq, and the Caucasus for the origin of this G2a branch during the Neolithic period.

In the North Caucasus, Boulygina et al. (2020) analyzed two ancient samples from the Koban culture. One of them belonged to G2a1-FGC1159, a sub-branch of G2a-Z6653, which is also present in Mingrelians from this study and has been dated to 5,414 ybp (https://www.yfull.com/tree/G-FGC1159/). In addition, an Alan male dating to the fifth to sixth century CE belonged to G2a-Z6653, while a Sarmatian individual belonged to the G2a-FGC1053 sub-branch of G2a-Z6653 (Gnecchi-Ruscone et al. 2022), which dates to 3,768 ybp (https://www.yfull.com/tree/G-FGC1053/). Due to the lack of SNP data from the interior branches of the G2a phylogeny, we could not confirm whether this subclade was represented in our Mingrelian samples.

Outside of the Caucasus, NRY haplotypes from G2a-Z6653 have been found in Europe. A sample from the Saltovo-Mayaki culture in the Belgorod region of Russia (Damgaard et al. 2018) likely belongs to G2a-Z6653, although the absence of terminal SNP (Neparáczki et al. 2018; Fóthi et al. 2020) prevents us from assessing its relationship with similar haplotypes in Mingrelians. Another sample from Hungary belonged to the G2a-FTB14662 sub-branch of G2a-Z6653 (Maróti et al. 2022). This sub-branch is found in various populations of the Caucasus including Mingrelians and is dated to 3,574 ybp (https://www.yfull.com/tree/G-FTB14662/). Thus, multiple subclades of G2a are present in the South Caucasus, suggesting this hg arose and diversified there prior to spreading to surrounding regions.

hg J2a

In the current study of 372 Mingrelian males, 10.75% of them belonged to hg J2a-Y11200 (J2a1a1a2b2a3b1a) (supplementary fig. S11, Supplementary Material online). Y-chromosomes belonging to this branch have a TMRCA of 6,980 ybp (https://www.yfull.com/tree/J-Y11200/). The oldest archaeological samples having J2a-Y11200 Y-chromosomes were found in the Northwest Caucasus among individuals from the Eneolithic Darkveti-Meshoko culture (4,700–3,500 cal BCE) (Trifonov 2009) and slightly later in the same region at sites associated with the Maykop Novosvobodnaya culture (fourth millennium BCE) (Nedoluzhko et al. 2014; Wang et al. 2019). The dominant subclade of J2a-Y11200 in Mingrelians was J2a-Y30811, which has been dated to 4,700 ybp (https://www.yfull.com/tree/J-Y30811/) (supplementary table S2, Supplementary Material online).

An additional 10.5% of Mingrelian males belonged to hg J2a-Y12379 (J2a1b2) (https://www.yfull.com/tree/J-Y12379/) and specifically to subclade J2a-Y12378 (https://www.yfull.com/tree/J-Y12378/). Y-chromosomes belonging to this branch have a TMRCA of 11,015 ybp (https://www.yfull.com/tree/J-Y12378/). The oldest ancient samples having J2a-Y12379 Y-chromosomes have been found at the Kotias Klde rock shelter site in western Georgia, which has been dated to 9,529–9,895 ybp (Jones et al. 2015). The most common sub-branch of J2a-Y12379 in Mingrelians is J2a-Y27964, which has been dated to 9,017 ybp (https://www.yfull.com/tree/J-Y27964/). Notably, J-Z6046, a branch of J2a1 ancestral to J-Y12379 (i.e., the haplotype of the Kotias Klde Mesolithic individual) (https://www.yfull.com/tree/J2/), is found in abundance among contemporary Mingrelians.

hg L

We found that approximately 7% of Mingrelian men had hg L1b (L-M317) Y-chromosomes, with none occurring in Svans (Yardumian et al. 2017) and only 1.5% of Georgians having them (Battaglia et al. 2009). These results are generally consistent with data appearing in the Georgian DNA Project, which indicates that Laz from Georgia and Turkey, Mingrelians, Georgians, and one Svan also belong to this paternal lineage (https://www.familytreedna.com/public/georgia/default.aspx?section=yresults). L1b also appears in the North Caucasus, albeit at low percentages (Balanovsky et al. 2011; Yunusbayev et al. 2012). In addition, hg L1b is also quite common among Pontic Greeks, being widespread in the southeastern Pontic region (Pontic and Anatolian Greek DNA Project; (https://www.familytreedna.com/public/russiangreeks?iframe=yresults).

Besides these regions, hg L1b is found among Lebanese Maronites (Platt et al. 2021), Italians, and Greeks (Battaglia et al. 2009) and populations from Iran, Afghanistan, and Pakistan (Di Cristofaro et al. 2013). Interestingly, the L1b haplotypes of the Maronites belong to a branch that "split from the Caucasus ancestral group around 7,300 years ago and migrated to the Levant" (Platt et al. 2021: 581). By contrast, the phylogeographic relationship between the Iranian and South Asian branches of L1b and those in Caucasus/Pontic group is unclear. Intriguingly, no L1b haplotypes are yet known from the aDNA record.

hg R1a

In this study, Mingrelians had 4.7% hg R1a-Z93 haplotypes, while both Svans (9.7%) and Georgians (10.6%) had higher frequencies of this paternal lineage. R1a-Z93 haplotypes have also been noted in Georgian individuals hailing from various parts of the country based on data from the Georgian DNA Project. In addition, R1a occurred in North Caucasus populations albeit at varying frequencies in different ethnic groups (supplementary table S2, Supplementary Material online).

Y-chromosomes belonging to this SNP hg date to 4,501 ybp (https://www.yfull.com/tree/R-Z93/). When assessing their age, the haplotypes of R1a-Z93 in Mingrelians gave a lower TMRCA of 2,930 ± 359 ybp. This date could possibly reflect a recent migration into the Caucasus or indicate a recent bottleneck of R1a-Z93 bearers in Mingrelians that led to reduced haplotypic diversity in this group. It should be noted, in this context, that the R1a-Z93 haplotypes present in the Srubnaya archaeological culture of the Ponto-Caspian steppe, which is thought to derive from the Yamanya culture, has been dated to 1,900–1,200 BCE (Engovatova et al. 2023; Saag et al. 2021).

hg R1b

We observed that 4.7% of Mingrelian men had hg R1b (R-Z2103/CTS1078) Y-chromosomes, with none occurring in Svans (Yardumian et al. 2017) and 9.1% of Georgians sampled by Battaglia et al. (2009) having them. These results are again consistent with data appearing in the Georgian DNA Project database, where Georgians from various parts of the country possess this paternal lineage. R1b is also present in North Caucasus populations at variable frequencies (supplementary table S2, Supplementary Material online), suggesting its wider presence in this region.

Interestingly, all 17 haplotypes from hg R1b in Mingrelians belonged to the R1b-Z2103 subclade, which also occurs frequently in Georgians. Y-chromosomes belonging to this subclade are estimated to have arisen 5,894 ybp (https://www.yfull.com/tree/R/). This subclade also occurs in 17 individuals from the Yamnaya culture (5,399–4,600 ybp) (Haak et al. 2015), which is proposed to have spread Indo-European languages into different parts of Eurasia (Anthony 2007; Reich 2018). In light of this evidence, it is possible that Yamnaya people spread R1b Y-chromosomes into the Caucasus region, where they now appear in modern-day Mingrelians. However, R1b-Z2103 haplotypes appear in numerous European and Middle Eastern populations, suggesting that there could possibly be other geographic sources of R1b haplotypes in Mingrelians.

Autosomal SNP Diversity in Mingrelians

We analyzed Mingrelian autosomal data in several different ways to elucidate their population history. The first was to project Mingrelians onto European or Eurasian populations that were previously genotyped on the Human Origins (HO) array in a principal component (PC) plot (Lazaridis et al. 2014) (fig. 4). Almost all Mingrelian individuals clustered together and were very similar to previously analyzed Mingrelians (Lazaridis et al. 2014) (red points in fig. 4). We also noted a general North Caucasus versus South Caucasus difference (c.f. Wang et al. 2019), as well as a distinction where some North Caucasus populations located in the piedmont (e.g., Chechens and Adygei) had Russian-like ancestry (but not necessarily ethnic Russian ancestry per se).

As previously noted, 20 individuals in our data set appeared as outliers in the PC analysis (PCA) plot (green points in fig. 4) and were excluded from further analysis. For most of them, individuals of Russian, Belorussian, or Ukrainian ancestry had married into their families over the past three generations, although the genealogical data for a small subset were not entirely clear on this point. Interestingly, none of the outliers included anyone having Abkhaz, Armenian, Georgian, Ossetian, or Svan ancestry in their family history unless they also included Slavic admixture. This result affirmed that we

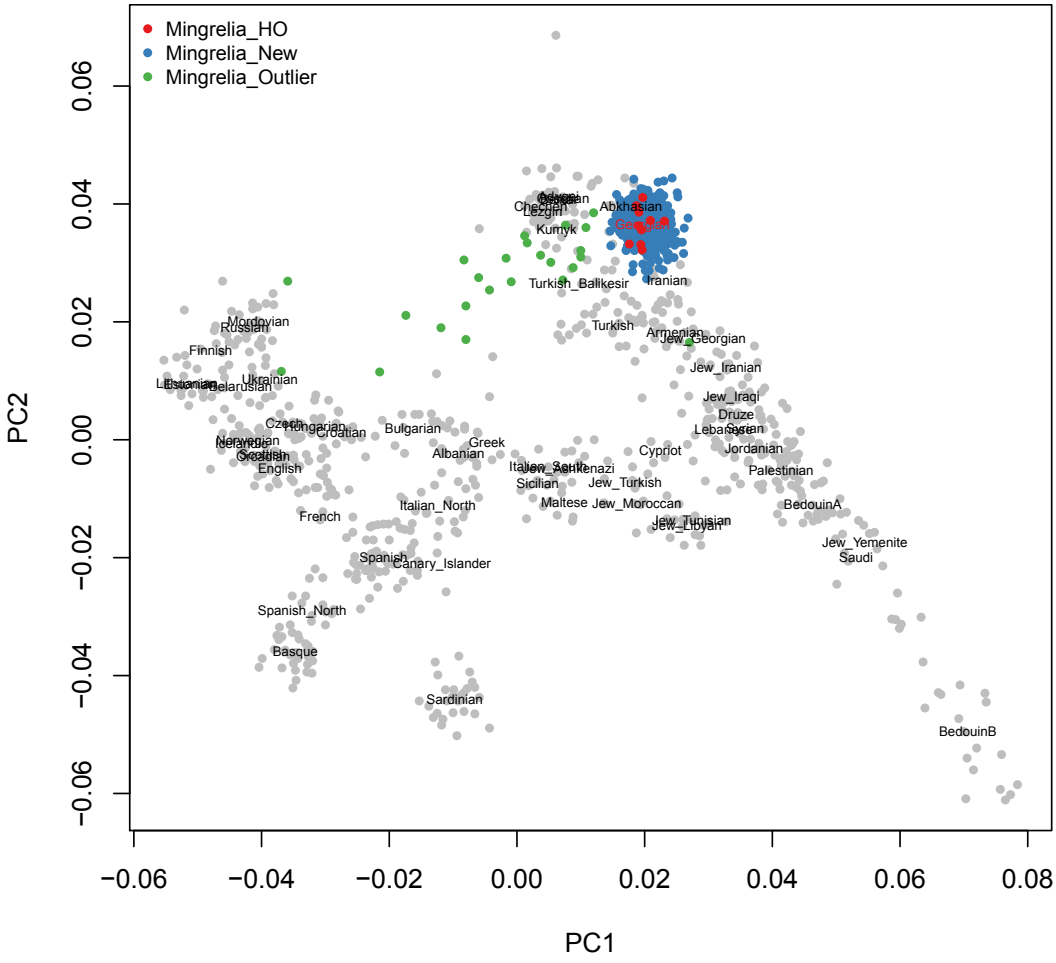

Fig. 4.—Mingrelian samples projected onto PCs defined by 777 West Eurasian individuals genotyped on the HO array. Mingrelian samples are shown in blue, whereas the 777 West Eurasian individuals genotyped on the HO array are shown as labeled gray points. Georgian individuals in the HO data (in fact Mingrelians) are shown in red. Outliers for Mingrelians, shown as green points, were removed from further analysis. According to Lazaridis et al. (2014), from which the comparative data are taken, PC1 and PC2 explain 0.9% and 0.4% of variance in this plot, respectively.

were detecting a Caucasus-based genomic signature in the Mingrelian samples on the PCA plot.

Within the main cluster of Mingrelian individuals, we found no evidence of population structure using PCA (supplementary fig. S12, Supplementary Material online) or supervised clustering with ADMIXTURE (supplementary fig. S13, Supplementary Material online). We did find some evidence of decreasing identity by descent (IBD) with distance between pairs of sampling locations (by about 0.1 Mb/km; supplementary fig. S14, Supplementary Material online), although this trend was not statistically significant (permutation *P*-value of 0.07, excluding the outlier pair of Taleri vs. Salkhino).

We found limited evidence for recent consanguinity. One individual had 3 long (>10) runs of homozygosity (ROH) and around 110 Mb in total, almost twice the mean and corresponding to parents who were second cousins (supplementary fig. S15, Supplementary Material online). Otherwise, the average total ROH was around 65 Mb. Based on estimated IBD, we identified 20 pairs of individuals with recent relatedness, including nine pairs of first-degree relatives, four pairs of second-degree relatives, and seven pairs of third-degree relatives. For the sake of clarity of outcome, we removed one individual from each of these pairs for the PCA, ADMIXTURE, and IBD analyses.

We further analyzed the Mingrelian autosomal data in relation to genomic data for only populations from the Caucasus, Turkey, Iran, and Russia to evaluate their genetic affinities on a more localized geographic scale (fig. 5*a*). Nearly all Mingrelian individuals in our study clustered together and were similar to the Mingrelians and Abkhazians genotyped by Lazaridis et al. (2014). Turkish, Armenian, and Iranian individuals also clustered together, as did most of the North Caucasus populations. As before, Russians were distinct from all other populations analyzed in this PCA plot.

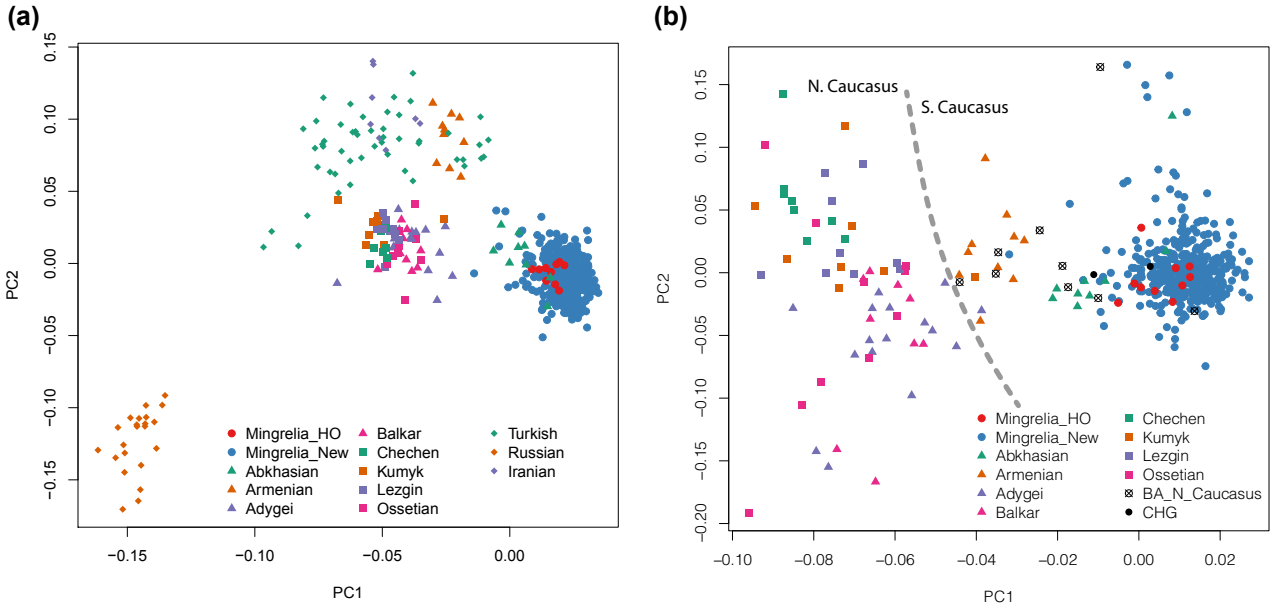

Fig. 5.—(*a*) PCs plot of newly genotyped Mingrelians and 171 individuals from the Caucasus, Turkey, Iran, and Russia genotyped on the HO array. PC1 and PC2 explain 0.6% and 0.3% of the variance in the plot, respectively. (*b*) PC analysis of newly genotyped Mingrelians and 91 individuals from present-day Caucasus populations genotyped on the HO array, with ancient Bronze Age North Caucasus individuals projected onto them. The black dots identify the CHG samples. The dashed line indicates the approximate divide between present-day North and South Caucasus populations. The black squares with an "x" inside of them represent data from Bronze Age individuals analyzed in Wang et al. (2019). The nine ancient individuals, the archaeological sites at which they were recovered, and their associated archaeological cultures include the following: ARM001 = Kaps site, Armenia, Kura-Araxes culture; ARM002 = Kaps site, Armenia, Kura-Araxes culture; I6266 = Klady site, (Adygea) Russia, Maykop Novosvobodnaya culture; I6267 = Klady site, (Adygea) Russia, Maykop Novosvobodnaya culture; I6268 = Klady site, (Adygea) Russia, Maykop Novosvobodnaya culture; I6272 = Klady/Dlinyana Polyana site, (Adygea) Russia, Maykop Novosvobodnaya culture; MK5004 = Mar'inskaya site, (Stavropol) Russia, Late Maykop culture; MK5008 = Mar'inskaya site, (Stavropol) Russia, Late Maykop culture; VEK007 = Velikent site, (Daghestan) Russia; Kura-Araxes culture. All of these aDNA samples have been dated to between 5,600 and 4,500 cal BP. PC1 and PC2 explain 0.4% and 0.3% of the variance in this plot, respectively (excluding the projected samples).

In addition to analysis with present-day populations, we compared the Mingrelian genomic data with those from ancient individuals from the Bronze Age Caucasus (Caucasus_BA) and the steppe region to the North (Wang et al. 2019) (fig. 5*b*). The dashed line shows the approximate geographic boundary between present-day North Caucasus and South Caucasus populations. Interestingly, the Bronze Age North Caucasus samples (Maykop and Kura-Araxes) projected onto present-day South Caucasus populations, while present-day North Caucasus populations were shifted to the left in the direction of Bronze Age steppe ancestry. The Bronze Age North Caucasus populations were also located just north of the mountains (on the left of the dotted line). This distinction was also shown by the *D* statistic *D*(Mbuti, Mingrelia, Caucasus_BA, Steppe Maykop) with $Z = -8.5$.

These results may reflect an expansion of steppe ancestry into the North Caucasus that did not extend south of the mountains. Without additional Bronze Age Georgian samples for comparison, though, it is difficult to make a firm conclusion about this pattern. Taken together, these results indicate a relatively genetically homogeneous population in Mingrelia and a significant degree of population continuity in the wider Caucasus region since at least since the Bronze Age.

Table 3.
D Statistics for Mingrelian and Comparative Populations in the Form *D*(Mbuti, X, Caucasus_BA, Mingrelia)

X	*D* (*Z* score)
Armenian	0.934
Sardinian	1.328
Russian	−1.765
Turks	0.582
North Ossetian	0.217
Iranian	−0.978
Han Chinese	−1.400
Georgian	2.872

Note.—For this standard *Z* score, a two-sided cutoff for $P = 0.05$ is ±1.96.

We further investigated whether the Mingrelian samples represented a clade with the Bronze Age Caucasus relative to neighboring present-day populations from the Simons Genome Diversity Project (Mallick et al. 2016) using *D* statistics of the form *D*(Mbuti, X, Caucasus_BA, Mingrelia) (table 3). We found no significant results except when X was a sample from present-day Mingrelia.

To assess possible links between Mesolithic and modern-day populations from the Caucasus, we projected genomic data

from the Kotias Klde and Satsurblia Caucasus Hunter-Gatherer (CHG) samples (Jones et al. 2015) onto the PCA plot of ancient and modern data (fig. 5*b*). Intriguingly, the CHG samples projected in the middle of a cluster containing ancient and modern South Caucasus populations. While tempting to interpret these results as indicating genetic continuity between the ancient CHG and present-day South Caucasus populations, they more likely reflect the fact that all modern populations share the same relative proportion of CHG ancestry (~30% CHG and ~70% Anatolian Chalcolithic ancestry; Wang et al. 2019) and are genetically equidistant from the CHG samples. The slight divergence of the North and South Caucasus clusters likely reflects later genetic "dilution" of North Caucasus groups from Steppe/European Hunter-Gatherer-related populations (Wang et al. 2019).

Discussion

This study presents genetic data from the largest number of Mingrelian (Georgian) individuals analyzed to date. Mingrelians show significant mtDNA and Y-chromosome diversity in terms of hg composition and haplotype diversity and are more diverse than Svans from the coextensive region north of them. This difference may be partly due to the Mingrelian sample being twice the size of the Svans but perhaps also the role that this region of western Georgia played in population movements through the South Caucasus. The presence of rare East Eurasian haplotypes in Mingrelians further points to complex stories to explain their presence in the South Caucasus.

Aside from the recent merger of Samegrelo and Upper (Zemo) Svaneti into a single administrative unit (fig. 1), these neighboring regions are likely to have been in commercial and perhaps bride-exchange contact for centuries. Historical accounts indicate that, for centuries, many great Svan men lived by performing migrant labor, dwelling almost full time in Samegrelo (Gelovani 2003). Still, differing frequencies of paternal hgs, including hgs L1b, Q, and R1b in Mingrelians, showed that these populations are not uniform in their genetic makeup.

Autosomal SNP data reveal Mingrelians to be genetically similar to Georgian individuals previously genotyped for autosomal markers. Mingrelians did not show significant genetic structure in the region, although there was tentative evidence of decreasing IBD with distance between pairs of sampling locations. In addition, the comparison of genomic data from Mingrelians and Bronze Age populations from the Caucasus revealed them to be largely the same. Taken together, these results indicate the presence of a relatively homogeneous population in Mingrelia, one that likely represents a high degree of population continuity in the wider Caucasus region since at least the Bronze Age, as further suggested by *D* statistics. These results further suggest that Mingrelians as a population may have evolved along with the Mingrelian language over this same period.

Along similar lines, our data suggest that the genetic profile of Mingrelians took shape form before the Colchian culture emerged in western Georgia circa 1,500 BCE (Apakidze 2008). In this regard, researchers have noted a continuity in material culture in Mingrelia through the Bronze and Iron Ages (e.g., Sagona 2017), by which time Svaneti to the north was likely settled (Chartolani 1974). Interestingly, despite this continuity, there is a lack of population structure in Samegrelo. This finding is generally consistent with our observation that Mingrelian surnames are not linked to specific Y-chromosome hgs, thus suggesting the absence of clans or lineages based on patrilineal affiliation.

If Mingrelian population diversity began taking shape around the Bronze Age of the Caucasus, then it becomes important to determine the groups that contributed to it. Among these groups could have been settlers from the greater Near East or local CHG and other past populations (genomic groups) who also influenced the pattern of genetic diversity in Eurasia. In this regard, Bronze Age South Caucasus populations have previously been modeled as having about 30% CHG-related ancestry, with the rest coming from populations related to Chalcolithic groups from present-day Iran and Armenian (Wang et al. 2019). Since our data show contemporary Mingrelians to be fairly similar to these Bronze Age populations, we expect their CHG ancestry proportions to be similar in composition. This interpretation is further supported by the *D* statistic for *D*(Mbuti, CHG, Caucasus_BA, Mingrelia), which is not significantly nonzero ($Z = 0.895$).

Still, these statistical results should be taken with some caution due to the small size of the CHG sample (two individuals) being used to generate them (Wang et al. 2019). Interestingly, the two CHG samples have mtDNA and Y-chromosome hgs commonly observed in the South Caucasus and Near East. Kotias Klde (KK1), which dates to 7,940–7,600 cal BP, has mtDNA hg H13c and NRY hg J2-Y12379*, whereas Satsurblia (SATP), which dates to 11,430–11,180 cal BP, has mtDNA K3 and NRY hg J1-FT34521 (Jones et al. 2015). Each of these hgs dates back to at least the Mesolithic (https://www.yfull.com/mtree/H13c/; https://www.yfull.com/tree/J-Y12379/; https://www.yfull.com/mtree/K3/; https://www.yfull.com/tree/J-FT34521/). One Neolithic Anatolian also has a hg J2a Y-chromosome and bears significant CHG ancestry (Mathieson et al. 2015), suggesting that CHG was also present in Neolithic Anatolia. What CHG individuals lack are hgs occurring at significant frequencies in modern-day Mingrelians, such as mtDNA X2 and U4 and NRY G2a. By contrast, other Bronze Age Caucasus samples show the same types of mtDNA and Y-chromosome hgs seen in contemporary Mingrelians, Svans, and Georgians (Wang et al. 2019). This observation suggests that the observed pattern of haploid genetic diversity had emerged by this time.

While present at low frequencies, the presence of eastern Eurasian mtDNA hgs dsC4a1a, D4, and F1b1 in an otherwise genomic western Eurasian population is curious and requires explanation. First, it is extremely unlikely that these lineages are survivors of Paleolithic settlement of Eurasia when the human genetic landscape may have looked vastly different. Instead, they likely reflect more recent population events, such as the westward expansions of Turkic-speaking populations. Evidence of these expansions is the presence of Turkic-speaking groups in the North Caucasus, such as Kumyks, Karachay-Balkars, and Nogais, and Azeris in the South Caucasus. There is also the Meskhetian dialect of Turkish, which arose from the regions of Kars, Ardahan, and Artvin and spread into nearby Samtskhe-Javakheti relatively recently (16th–17th century CE) (Aydıngün et al. 2006). Although individuals speaking the latter two languages have not yet been adequately sampled for mtDNA diversity, a comparison of data from North Caucasus Turkic-speaking peoples (e.g., Bermisheva et al. 2004; Yunusbayev et al. 2012) with those from Mingrelians suggests no specific connections between them.

Given this evidence, it may be the case that the East Eurasian mtDNA hgs detected among Mingrelians (and to a lesser extent Svans) are vestiges of some settlement event associated with the Golden Horde (Suny 1994). A possible source for them is the Ilkhanate, an Iran-based Mongol imperial realm that briefly subsumed the Middle East, including parts of the South Caucasus (Lane 2022). Indeed, the Mongols defeated Georgian armies and occupied the region throughout the 13th century until withdrawing from it in the early 14th century (Suny 1994). Whatever their link to this history of conquest and resistance, these haploid lineages have their strongest focus in Central and East Asia, indicating that their source populations derived from these regions or from areas influenced by population expansions emanating from the east.

When viewed from a broader lens, Mingrelians and Georgians show genetic similarities to non-Kartvelian-speaking populations occupying the regions around them. Our previous research in Georgia indicated that geography was a stronger predictor of genetic relationships in the Caucasus than linguistic affiliation. We observed that speakers of Svan, Abkhaz, and Ossetian, which all belong to different language families, showed a fairly tight clustering in PCA plots based on mtDNA and Y-chromosome data (Yardumian et al. 2017). The results from our current study largely confirm this trend, although revealing subtle genetic differences between North and South Caucasus populations based on autosomal data, as well as some between North and South Ossetians. Moreover, the fact that groups that are sometimes associated with the South Caucasus (e.g., Armenians) align with Turkish and Iranian populations is consistent with most ancient DNA studies, which show genetic continuity between Anatolia and the South Caucasus/Zagros Mountain region (e.g., Skourtanioti et al. 2020).

An apparent exception to this pattern is the clustering of speakers of Mingrelian (Kartvelian) and Abkhaz (Northwest Caucasian) languages in the PCA plots. This result should perhaps not be surprising because Mingrelians and Abkhazians lived in overlapping regions of the western and northwestern Caucasus until very recently (Chervonnaya 1994). Additional data from South Ossetian individuals may further add to our understanding of genetic variation in the regions of Georgia located closest to the North Caucasus Mountains.

Adding complexity to the interpretation of these genetic data are population movements into and shifts within the Caucasus during the historical period. As an example, there was an influx of Armenians into present-day Armenia in the wake of the early 20th-century Armenian Genocide in Turkey (Hovannisian 1971), and many North Caucasus peoples left the Russian Empire and settled in Turkey during the late 19th century (Chochiev 2007). In addition, potential genetic influences from Azerbaijan are unclear because Azeris are not especially well represented in anthropological genetic studies. Thus, more work is needed to clarify the potentially cultural and genetic contributions of these groups to Georgians.

From a genealogical perspective, our autosomal data further indicate that the Slavic influence on persons with Georgian and Mingrelian ancestry is recent (within several generations) and came through both males and females. This influence is clearly observed in the admixture data for Mingrelians, where individuals with some Slavic ancestry fall outside the main cluster of Mingrelians in the PCA plots based on autosomal data. Thus, Slavic populations are clearly genetically distinctive from South Caucasus groups.

In the future, we will explore the patterns of genetic variation observed in the mtDNA and NRY data sets in several ways. This work will include mitogenome sequencing and Y-chromosome resequencing of selected samples from key hgs present in the region (e.g., mtDNA U4b1b and NRY G2a). The resulting data will likely confirm the presence of unique lineages or haplotypes in western Georgia and provide in-depth sequence data that can be used in coalescence analysis to estimate the ages of these hg and their subclades in the region. The data will also expand our understanding of the broader phylogenetic history of the mtDNA and NRY hgs present in Mingrelians. Autosomal and uniparental data from additional western Georgian populations will further reveal whether the pattern of genetic diversity observed in Samegrelo is consistent for all regions of Georgia or instead reveal east–west differences mirroring the historical regions of Colchis and Iberia.

Materials and Methods

Sample and Data Collection

In August 2016, a joint American-Georgian research team conducted fieldwork in 29 different cities, towns, and villages

in Samegrelo (supplementary fig. S16, Supplementary Material online). These locations included Abasha, Akhalsopeli, Chkhorotsqu, Didi Chqoni, Jbali, Jvari, Khabume, Kheta, Khobi, Kurzu, Letsurtsume, Lia, Martvili, Mukhuri, Nojikhevi, Nokalakevi, Norio, Nosiri, Obuji, Poti, Potskho-Etseri, Salkhino, Senaki, Sujuna, Taleri, Teklati, Tsalenjikha, Zana, and Zugdidi. We also consented a small number of Georgian and Mingrelian participants in Tbilisi. The consent and sampling protocols, as well as the procedures used for DNA analysis, were conducted with the permission of the University of Pennsylvania IRB #8 and the Georgian National Council on Bioethics.

Written informed consent was obtained from participants using a Georgian language form prior to the collection of DNA samples with buccal swabs. As part of the enrollment process, genealogical information was also collected from all participants. Participants were asked to provide their age and birthplace, their parents' names, ethnicity, and birthplaces, and similar information for their four grandparents. This information, along with extended genealogical interviews conducted with a subset of participants, yielded important details about the demography of Samegrelo. Although all unrelated males and females were encouraged to participate, when working with men, emphasis was placed on enrolling individuals having different Mingrelian surnames (i.e., patrilineal clan-affiliated lineages).

Through this approach, we enrolled a total of 485 participants in the study, including 372 males and 113 females. Of this number, 476 were Mingrelian speakers, six were non-Mingrelian-speaking Georgians (females: SAM 478 and 485; males: SAM 479, 480, 481, and 483), one individual was a male Georgian citizen without any Georgian genetic ancestry (SAM 474), and two individuals were females from Abkhazia with no Georgian genetic ancestry (SAM 075 and 076). We did not include the data for these nine individuals in the totals for Mingrelian speakers. These exclusions left a total of 476 individuals for the mtDNA analysis and 367 males for the NRY analysis of Mingrelian speakers.

Of the 476 Mingrelian speakers, 14 reported having a non-Mingrelian-speaking maternal mother or grandmother who was either Belorussian, Russian, or Ukrainian (SAM 055, 065, 071, 097, 099, 136, 149, 258, 343, 372, 419, 423, 443, and 475), while another person reported having a non-Mingrelian-speaking father who was Russian (SAM 366). The data for these admixed Mingrelian-speaking individuals are reported in the supplementary tables but were not used for subsequent analyses of Mingrelian genetic diversity. These 14 exclusions left 462 individuals for mtDNA analysis and 366 males for NRY analysis, respectively. We discuss the process of further exclusions of individuals based on autosomal data below.

During fieldwork, we attempted to enroll participants who did not share any consanguineal relationship with other participants through three generations. However, upon subsequent review of the participants' genealogical data, we noted that some relatives did, in fact, participate in the study. They included 1) a pair of brothers with the same father and mother (SAM 293 + SAM 294); 2) two pairs of brothers and sisters who had the same parents (SAM 131 + SAM 148; SAM 206 + SAM 219); 3) one father and son (SAM 458 + SAM 464); 4) one mother and daughter (SAM 462 + SAM 465); 5) three mothers and sons (SAM 037 + SAM 038, SAM 133 + SAM 134, and SAM 312 + SAM 316); and 6) one father and daughter (SAM 430 + SAM 432).

Based on this information, we excluded 1) one of the brothers for both mtDNA and Y-chromosome analyses in the first case (SAM 293); 2) excluded the brothers from the mtDNA analysis in the second (SAM 131 and SAM 219); 3) excluded the son from the NRY analysis in the third (SAM 458); 4) excluded the daughter for mtDNA analysis in the fourth (SAM 465); 5) excluded the three sons from mtDNA analysis in the fifth (SAM 037, SAM 133, and SAM 316); and 7) kept both individuals for mtDNA analysis in the sixth. In addition, one sample did not yield any usable mtDNA sequence data for analysis (SAM 254). These final eight exclusions left a total of 454 individuals for the further analysis of mtDNA diversity and 364 males for the analysis of Y-chromosome diversity in Mingrelians. The mtDNA and NRY data for the relatives are reported in supplemental tables 1 and 2, Supplementary Material online.

Genomic DNA Preparation

Genomic DNAs were extracted from buccal swab samples using Puregene extraction kits (Qiagen), according to the manufacturer's protocol.

mtDNA Analysis

For all 485 samples (irrespective of the exclusions), the entire mtDNA CR, which encompasses hypervariable regions 1 and 2 (HVS1 and HVS2), was PCR amplified and sequenced using published methods (Schurr et al. 2012). All polymorphic nucleotides were reckoned relative to the revised Cambridge Reference Sequence (Anderson et al. 1981; Andrews et al. 1999) and checked against the Reconstructed Sapiens Reference Sequence (Behar et al. 2012). The CR sequence data defined maternal haplotypes in these individuals. The hg status of each haplotype was ascertained with *haplogrep classify* v2.2.8 (Kloss-Brandstatter et al. 2011) and checked against Phylotree Build 17 (van Oven and Kayser 2009).

Y-Chromosome DNA Analysis

Sequence variation in the NRY for the 372 male participants was characterized through the analysis of 17 Y-STRs in the ABI AmpFLSTR Yfiler PCR Amplification Kit, as previously

described (Schurr et al. 2012). These markers were run on a 3730xl Genetic Analyzer and read with GeneMapper ID v3.2 software. The combination of alleles at these loci defined the Y-STR haplotype for each male individual. The hg status of all Y-STR haplotypes was determined using the Nevgen Y-DNA Haplogroup Predictor (http://www.nevgen.org/), with the results being compared with the ISOGG 2016.01.04 tree, which is based on Y-chromosome sequence data.

Autosomal SNP Genotype Analysis

All 485 individuals were genotyped at 645,337 SNPs on the GenoChip 2.0+ microarray at Gene-by-Gene Ltd (Houston, TX, USA). All microarrays were run on the Illumina iScan platform and data were processed using Illumina's GenomeStudio software. Data files for each individual in .csv format were converted to .ped format for further analysis.

Using *plink* (Purcell, et al. 2007), we removed 27,011 SNPs with greater than 5% missingness and excluded 21 individuals having greater than 2% missingness in their SNP data. These individuals included SAM 006, 024, 039, 048, 068, 120, 123, 148, 154, 167, 186, 211, 225, 230, 233, 252, 254, 264, 288, 438, and 467. Their exclusion left 464 individuals for further autosomal analysis.

We merged these genotypes with the HO data (57,566 SNPs in the intersection) and projected the Mingrelian samples onto the PCs defined by 777 West Eurasian individuals (Lazaridis et al. 2014) using *smartpca* (Patterson et al. 2012). Through this process, we identified 25 individuals who were outliers relative to other Mingrelians on the resulting PC plot and removed them from further analysis of autosomal SNP genotypes. They included SAM 016, 024, 055, 065, 071, 075, 076, 097, 099, 136, 149, 159, 161, 258, 265, 276, 281, 343, 359, 366, 372, 432, 443, 454, 474, and 475. These exclusions further reduced the number of individuals for focused autosomal analysis to 439.

We additionally used *plink --homozyg* to identify long ROH using default parameters; that is, each ROH had to contain at least 100 SNPs and be longer than 1 Mb, in order to assess relatedness between pairs of individuals. We also used *plink --genome* to identify 20 pairs of individuals having relatedness of >0.09375 (i.e., between first- and third-degree relatives) based on estimated IBD sharing and removed one from each pair. Those removed included SAM 018, 038, 060, 133, 150, 181, 191, 202, 206, 253, 294, 295, 303, 316, 331, 330, 424, 432, 458, and 462. As a result of these steps, the final autosomal SNP data set consisted of 419 individuals genotyped at 618,326 SNPs.

Having generated this refined Mingrelian SNP data set, we computed PC using *smartpca* (Patterson et al. 2012) and performed unsupervised clustering using *ADMIXTURE* (Alexander et al. 2009) with $K = 2$ to $K = 6$. We phased the data using *beagle4* (Browning and Browning 2007) and then used *refined-ibd* (Browning and Browning 2013) to identify IBD segments in the phased data.

In addition, we extracted 5,205 mtDNA and 10,272 Y-chromosome SNPs from the array data and used *haplogrep classify* (v2.2.8) (Kloss-Brandstatter et al. 2011) and *yhaplo* (v1.1.0) (Poznik 2016) with the ISOGG 2016.01.04 tree to call mtDNA and NRY hgs, respectively. The hg calls were subsequently compared with those previously generated from mtDNA CR sequences and Y-STR haplotypes, as described above, to evaluate their commensurability. The mtDNA and Y-chromosome SNP haplotypes and their associated hgs are reported in supplementary tables S1 and S2, Supplementary Material online, respectively.

Comparative Genomic Data Sets

To determine the phylogeographic affinities of Mingrelians, we compared their genetic data to those of populations from Europe, Caucasus, and the Middle East. Both mtDNA and NRY data were drawn from published data for Abkhazians, Armenians, Circassians, Georgians, Svans, Kabardians, North Ossetians, South Ossetians, Anatolians, and Iranians (Nasidze et al. 2004; Balanovsky et al. 2011; Gökçumen et al. 2011; Herrera et al. 2012; Terreros et al. 2011; Yunusbayev et al. 2012; Yardumian et al. 2017).

For the autosomal analysis, we compared the Mingrelian data to HO SNP array data reported by Lazaridis et al. (2014), whole-genome sequence data from the Simons Genome Diversity Project (Mallick et al 2016), and ancient individuals genotyped using the 1240k capture reagent (Wang et al. 2019). We computed PCs for Mingrelian and comparative samples using *smartpca* (Patterson, et al. 2012). In some cases, we projected Mingrelian or ancient samples onto a PC plot defined by other sets of samples using the *lsqproject* option. Finally, we computed *D* statistics using *qpDstat* (Patterson, et al. 2012).

Statistical and Phylogenetic Analyses

To explore the phylogenetic history of the genetic lineages that were present in Mingrelians, we analyzed mtDNA HVS1 sequences with NETWORK 4.6.1.3 (Bandelt et al. 1999). All networks were visualized using Network Publisher v1.2.0.0 (Fluxus Technology). The mutation-weighting scheme was based on that described in Bandelt et al. (2002), in which fast-evolving sites were given lower weights relative to other less mutable sites. All variants known to result from homopolymeric C expansions (e.g., A16182C and A16183C) or to occur at mutational hotspots in the mtDNA CR (e.g., T16519C) were excluded from the haplotypes used in this analysis.

The hg status of all Y-STR haplotypes was determined using the Nevgen Y-DNA Haplogroup Predictor (http://www.nevgen.org/), which infers lineage status from the

ISOGG 2016.01.04 tree, which is based on Y-chromosome sequence data. These results were also compared with NRY hgs predicted from the microarray SNP analysis. Networks of Y-chromosomes were constructed using these SNP data.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We wish to express our deepest gratitude to the people of Samegrelo for their hospitality and participation in this study. Without their interest and involvement in this project, our study would not have been possible. We would also like to acknowledge the support and research assistance from the administrators, medical staff, historians, and journalists who helped facilitate our research in Samegrelo. They include Goga Gulordava (Tsalenjikha); Mamuka Kakachia (Obuji); Nanuli Badzaghua and Ether Beria (Potshko-Etseri); Murman Tsotsoria and Alexandre Grigalva (Martvili); Bachana Beselia (Kurzu); Malkhaz Nachqebia and Davit Todua (Taleri); Ilya Kukhakashvili (Salkhino); Drs. Gocha Lipartia and Jamal Gogokhia (Jvari); Merab Kvaratskhava (Jgali); Dato Gogua and Nona Papova (Chkhrototsqu); Nona Shengelia and Zurab Gulia (Khabume); Solomon Kosmava and Igor Tolordava (Mukhuri); Badri Sichinava (Letsurisume); Roland Gabadava (Akhuti); Gocha Dgebuadze, Mzia Maskhulia and Revaz Asmava (Senaki); Ia Abuladze (Teklati); Zurab Apakidze and Mikheil Khorava (Zana); Nino Gabelia and Vasil Beselia (Nosiri); Igor Baghaturia (Nokalakevi); Ekaterina Khuntsaria (Akhalsopeli); Davit Rusia (Zugdidi); Gocha Kajaia, Irina Turava, Gogita Chitaia, and Irina Pipilia (Khobi); Gocha Sichinava and Merab Chitaia (Nojikheji); Badri Chitanava (Kheta); Teimuraz Gabelaia; Mamuka Kvetashvili and Dmitri Tsanava (Abasha); Mamia Kutsia and Gela Khoshtaria (Sujuna); and Irakli Kakulia and Gela Chqonia (Poti). We are also grateful to Joni Anjaparidze, our driver who safely escorted us around Samegrelo during our fieldwork there and to Mikhail Shengelia for assisting with the photodocumentation of the expedition. We want to further acknowledge the organizations that have provided administrative and logistical support for this project. They include the University of Pennsylvania Museum of Archeology and Anthropology, the University of Pennsylvania Department of Anthropology, Tbilisi State Medical University, Ivane Javakhishvili Institute of History and Ethnology, and the Georgian Academy of Sciences. In addition, we would like to thank Dr. Giorgi Kavtaradze for valuable suggestions for and guidance with this project and Dr. Kevin Tuite from the Université de Montréal for his constructive reading of the manuscript. We are further grateful to Dr. Connie Bormans from Gene-by-Gene, Ltd., for her assistance with the microarray analysis. This work was supported by the National Science Foundation (BCS-1824826; T.G.S. and A.Y); the University of Pennsylvania Museum of Archaeology and Anthropology (T.G.S.); the University of Pennsylvania Faculty Research Funds (T.G.S.); the Paul Carpenter Scholars Program (A.Y.); and Bryn Athyn College Summer Research Funds (A.Y.).

Data Availability

The mtDNA CR sequence data for this study are reported in supplementary table S1, Supplementary Material online, while the Y-chromosome STR data for this study are reported in supplementary table S2, Supplementary Material online. The SNP data for this study have been deposited in Zenodo under DOI # 10.5281/zenodo.10045555.

Funding

This work was supported by National Science Foundation BCS-1824826 (T.G.S., A.Y); the University of Pennsylvania Museum of Archaeology and Anthropology (T.G.S.); the University of Pennsylvania Faculty Research Funds (T.G.S.); and the Bryn Athyn College Paul Carpenter Scholars Program and College Summer Research Funds (A.Y.).

Literature Cited

Agladze G, et al. 1998. Biological diversity of the Black Sea basin area: its current state. In: Kotlyakov V Uppenbrink VM and Metreveli V, editors. Conservation of the biological diversity as a prerequisite for sustainable development in the Black Sea Region. New York: Springer. p. 1–15.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19: 1655–1664.

Allentoft ME, et al. 2015. Population genomics of Bronze Age Eurasia. Nature 522:167–172.

Anderson S, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Andrews RM, et al. 1999. Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA. Nat Genet. 23:147.

Anthony DW. 2007. The horse, the wheel, and language: how bronze age riders from the eurasian steppes shaped the modern world. Princeton: Princeton University Press.

Apakidze J. 2008. Ancient Colchian pottery from Georgia. In: Rubinson KS and Sagona A, editors. Ceramics in transitions: Chalcolithic through Iron Age in the highlands of the Southern Caucasus and Anatolia. Leuven: Peeters. p. 323–368.

Aydıngün A, Harding CB, Hoover M, Kuznetsov I, Swerdlow S. 2006. Meshketian Turks: an Introduction to their history, culture and resettlement experiences. Washington (DC): Center for Applied Linguistics.

Balanovsky O, et al. 2011. Parallel evolution of genes and languages in the Caucasus region. Mol Biol Evol. 28(10):2905–2920.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet. 71:1150–1160.
Barbujani G, Nasidze I, Whitehead GN. 1994. Genetic diversity in the Caucasus. Hum Biol. 66(4):639–668.
Battaglia V, et al. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. Eur J Hum Genet. 17: 820–830.
Behar DM, et al. 2008. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3(4):e2062.
Behar DM, et al. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 90(4): 675–684.
Bermisheva M, et al. 2004. Phylogeographic analysis of mitochondrial DNA in the Nogays: a strong mixture of maternal lineages from eastern and western Eurasia. Mol Biol. 38(4):516–523.
Boeder W. 2005. The South Caucasian languages. Lingua 115:5–89.
Boulygina E, et al. 2020. Mitochondrial and Y-chromosome diversity of the prehistoric Koban culture of the North Caucasus. J Archaeol Sci: Rep. 31:102357.
Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 81:1084–1097.
Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194:459–471.
Chartolani ShG. 1974. Svaneta v Epokhu Bronzy [Svaneti in the Bronze Age]. PhD Dissertation. Tbilisi [in Russian].
Chervonnaya S. 1994. Conflict in the Caucasus: Georgia, Abkhazia and the Russian Shadow. Glastonbury: Gothic Image Publications.
Chirikba VA. 2008. The problem of the Caucasian Sprachbund. In: Muysken P, editors. From linguistic areas to areal linguistics. Philadelphia: John Benjamins. p. 25–94.
Chochiev G. 2007. On the history of the North Caucasian diaspora in Turkey. Iran Caucasus. 11:213–226.
Courcier A. 2014. Ancient metallurgy in the Caucasus from the sixth to the third millennium BCE. In: Roberts BW, Thornton CP, editors. Archaeometallurgy in global perspective: methods and syntheses. New York: Springer. p. 579–664.
Damgaard PDB, et al. 2018. 137 Ancient human genomes from across the Eurasian steppes. Nature 557:369–374.
Derenko M, et al. 2010. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in Northern Asia. PLoS One. 5(12): e15214.
Derenko M, et al. 2013. Complete mitochondrial DNA diversity in Iranians. PLoS One 8(11):e80673.
Derenko M, et al. 2014. Western Eurasian ancestry in modern Siberians based on mitogenomic data. BMC Evol Biol. 14:217.
Derenko M, et al. 2019. Insights into matrilineal genetic structure, differentiation and ancestry of Armenians based on complete mitogenome data. Mol Genet Genom. 294(6):1547–1559.
Derenko M, Denisova G, Malyarchuk B, Dambueva I, Bazarov B. 2018. Mitogenomic diversity and differentiation of the Buryats. J Hum Genet. 63(1):71–81.
Diakonoff IM, Starostin SA. 1986. Hurro-Urartian as an Eastern Caucasian language. München er Studien zur Sprachwissenschaft. Beiheft. n.F., 12. München: R. Kitzinger.
Di Cristofaro J, et al. 2013. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS One. 8(10):e76748.
Dulik MC, et al. 2012. Mitochondrial DNA and Y-chromosome variation provides evidence for a recent common ancestry between native Americans and indigenous Altaians. Am J Hum Genet. 90: 229–246.
Engovatova AV, et al. 2023. Ancient DNA of the bearers of the Fatyanovo and Abashevo cultures (concerning migrations of the Bronze Age people in the forest belt on the Russian Plain). Stratum Plus. 2:207–228.
Fernandes V, et al. 2012. The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. Am J Hum Genet. 90:347–355.
Fóthi E, et al. 2020. Genetic analysis of male Hungarian conquerors: European and Asian paternal lineages of the conquering Hungarian tribes. Archaeol Anthropol Sci. 12:31.
Gelovani A. 2003. Dasavlet Sakartvelos Mtianeti (Svaneti) XIX Saukuneshi [western Georgia's Highland Region (Svaneti) in the XIX century]. Tbilisi: Egrisi [in Georgian].
GeoStat. 2016. The 2014 general population census: main results. Tbilisi: National Statistics Office of Georgia.
Gnecchi-Ruscone GA, et al. 2022. Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites. Cell 18(8):1402–1413.e21.
Gökçumen Ö, et al. 2011. Biological ancestries, kinship connections and projected identities in four central Anatolian settlements. Am Anthropol. 113(1):116–131.
Grossheim AA. 1952. Plant wealth of the Caucasus [Rastitelnye Bogatstva Kavkaza]. Moscow: Moskovskoye Obshchestvo Ispytateley Prirody [in Russian].
Haak W, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 522:207–211.
Haber M, et al. 2016. Genetic evidence for an origin of the Armenians from bronze age mixing of multiple populations. Eur J Hum Genet. 24(6):931–936.
Herrera KJ, et al. 2012. Neolithic patrilineal signals indicate that the Armenian Plateau was repopulated by agriculturalists. Eur J Hum Genet. 20(3):313–320.
Hewitt G. 2017. Instilling pride by raising a language's Prestige. In: Korkmaz R, Doğan G, editors. Endangered languages of the Caucasus and beyond. Leiden: Brill. p. 91–97.
Hovannisian RG. 1971. The republic of Armenia. Vol. 1: the first year: 1918–1919. Berkeley: University of California Press.
International Society of Genetic Genealogy (ISOGG)—https://isogg.org.
Järve M, et al. 2019. Shifts in the genetic landscape of the Western Eurasian steppe associated with the beginning and end of the Scythian dominance. Curr Biol. 29:2430–2441.e10.
Javakhishvili I. 1998. Introduction to the History of the Georgian Nation, I: Historical-ethnological problems of Georgia, the Caucasus and the Near East. Essays, Vol. 10. Tbilisi: Tbilisi State University Publishing HouseVol. 10.
Jones ER, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun. 6:8912.
Juras A, et al. 2018. Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. Sci Rep. 8:11603.
Juras A, et al. 2021. Maternal genetic origin of the late and final Neolithic human populations from present-day Poland. Am J Biol Anthropol. 176(2):223–236.
Kassian A. 2010. Hattic as a Sino-Caucasian language. In: Dietrich M, Loretz O, editors. Ugarit-Forschungen. Internationales Jahrbuch für die Altertumskunde Syrien-Palästinas. Bd 41, 2009–2010. Münster: Ugarit-Verlag. p. 309–447.
Kavtaradze G. 1983. K Khronologii Epokhi Neolita i Bronzy Gruzii [On the Chronology of the Neolithic and Bronze Age of Georgia]. Tbilisi: Metsniereba [in Russian].
Kavtaradze G. 2002. An Attempt to Interpret Some Anatolian and Caucasian Ethnonyms of the Classical Sources. Sprache und Kultur #3 (Staatliche Ilia Tschawtschawadse Universitaet Tbilisi für Sprache und Kultur Institut zur Erforschung des westlichen Denkens), pp. 68–83.

Khusnutdinova EK, et al. 2012. Gene pool of ethnic groups of the Caucasus: results of integrated study of the Y chromosome and mitochondrial DNA and genome-wide data. Russ J Genet. 48(6):640–650.
Kılınç GM, et al. 2018. Investigating Holocene human population history in North Asia using ancient mitogenomes. Sci Rep. 8:8969.
Klimov GA. 1998. Etymological dictionary of the Kartvelian languages. Berlin: Mouton de Gruyter.
Kloss-Brandstatter A, et al. 2011. Haplogrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat. 32:25–32.
Kong QP, et al. 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet. 15:2076–2086.
Lane G. 2022. The ilkhanate. In: May T, Hope M, editors. The Mongol world. New York: Routledge. p. 279–297.
Lazaridis I, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413.
Lazaridis I, et al. 2016. Genomic insights into the origin of farming in the ancient near east. Nature 536:419–424.
Lazaridis I, et al. 2022. The genetic history of the Southern Arc: a bridge between West Asia and Europe. Science 377(6609): eabm4247.
Lipson M, et al. 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature 551(7680): 368–372.
Loo JH, et al. 2014. Mitochondrial DNA association study of type 2 diabetes with or without ischemic stroke in Taiwan. BMC Res Notes. 7:223.
Mallick S, et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201–206.
Malyarchuk B, et al. 2010. Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. Mol Biol Evol. 27(10):2220–2226.
Maróti Z, et al. 2022. The genetic origin of Huns, Avars, and conquering Hungarians. Curr Biol. 32:2858–2870.e7.
Mathieson I, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528(7583):499–503.
Mathieson I, et al. 2018. The genomic history of Southeastern Europe. Nature 555:197–203.
Metspalu M, Kivisild T, Bandelt HJ, Richards M, Villems R. 2006. The pioneer settlement of modern humans in Asia. In: Bandelt HJ, Macaulay V, Richards M, editors. Human mitochondrial DNA and the evolution of Homo sapiens. Heidelberg: Springer-Verlag. p. 181–199.
Mielnik-Sikorska M, et al. 2013. The history of Slavs inferred from complete mitochondrial genome sequences. PLoS One 8(1):e54360.
Narasimhan VM, et al. 2019. The formation of human populations in South and Central Asia. Science 365:eaat7487.
Nasidze I, et al. 2004. Mitochondrial DNA and Y-chromosome variation in the Caucasus. Ann Hum Genet. 68:205–221.
Nedoluzhko AV, et al. 2014. Analysis of the mitochondrial genome of a Novosvobodnaya culture representative using next-generation sequencing and its relation to the Funnel Beaker culture. Acta Naturae. 6(2):21–35.
Neparáczki E, et al. 2018. Mitogenomic data indicate admixture components of central-inner Asian and Srubnaya origin in the conquering Hungarians. PLoS One 13(11):e0208295.
Ning C, et al. 2020. Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat Commun. 11:2700.
Oštir K. 1921. Beiträge zur Alarodischen Sprachwissenschaft. Leipzig: Ed. Beyers Nachfolger.
Patterson N, et al. 2012. Ancient admixture in human history. Genetics 192:1065–1093.
Platt DE, et al. 2021. Autosomal genetics and Y-chromosome haplogroup L1b-M317 reveal Mount Lebanon Maronites as a persistently non-emigrating population. Eur J Hum Genet. 4: 581–592.
Poznik GD. 2016. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv. https://doi.org/10.1101/088716.
Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81: 559–575.
Reich D. 2018. Who we are and how we got here: ancient DNA and the new science of the human past. New York: Pantheon.
Reidla M, et al. 2003. Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet. 73(5):1178–1190.
Rivollat M, et al. 2020. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. Sci Adv. 6(22):eaaz5344.
Röck AW, Dür A, van Oven M, Parson W. 2013. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). Foren Sci Int Genet. 7:601–609.
Roostalu U, et al. 2006. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the near eastern and Caucasian perspective. Mol Biol Evol. 24(2): 436–448.
Saag L, et al. 2021. Genetic ancestry changes in Stone to Bronze Age transition in the East European Plain. Sci Adv. 7:eabd6535.
Sagona A. 2017. The archaeology of the Caucasus: from earliest settlements to the Iron Age. Cambridge: Cambridge University Press.
Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. 2011. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. Eur J Hum Genet. 19(9):988–994.
Schurr TG, et al. 2012. Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. Am J Phys Anthropol. 148(3):422–435.
Shlush LI, et al. 2008. The Druze: a population genetic refugium of the near east. PLoS One. 3(5):e2105.
Skourtanioti E, et al. 2020. Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. Cell 181(5): 1158–1175.
Suny RG. 1994. The making of the Georgian nation, 2nd ed. Bloomington: Indiana University Press.
Terreros M, Rowold D, Mirabal S, Herrera R. 2011. Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. J Hum Genet. 56:235–246.
Trifonov V. 2009. Was there the Neolithic in the Northern-Western Caucasus? [Sushchestvoval li na Severo-Zapadnom Kavkaze neolit?]. In: Trifonov VA, editors. Adaptaciya kultur Paleolita-Eneolita k izmeniyam prirodnoy sredy na Severo-Zapadnom Kavkaze [adaptation of the cultures of the Paleolithic-Eneolithic to the environment in the Northern-Western Caucasus]. Teza: Institut Istorii Materielnoy Kultury RAN, St. Petersburg. p. 84–93. [In Russian].
Vamling K, Tchantouria R. 2005. Mingrelian and Laz. http://www.academia.edu/2494004/Mingrelian_and_Laz
van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mut. 30:e386–e394.
Vasilyev S, Amirkhanov H. 2018. Palaeolithic Caucasus: paleoanthropological panorama. Quat Int. 465:105–116.
Wang HW, et al. 2012. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet. 57: 228–234.
Wang CC, et al. 2019. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat Commun. 10:590.

Yardumian A, et al. 2017. Genetic diversity in Svaneti and its implications for the human settlement of the Highland Caucasus. Am J Phys Anthropol. 164(4):837–852.
Yunusbayev B, et al. 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol Biol Evol. 29(1):359–365.
Zhang W, et al. 2014. A matrilineal genetic legacy from the last glacial maximum confers susceptibility to schizophrenia in Han Chinese. J Genet Genom. 41(7):397–407.
Zhang AM, et al. 2016. Mitochondrial DNA haplogroup A decreases the risk of drug addiction but conversely increases the risk of HIV-1 infection in Chinese addicts. Mol Neurobiol. 53(6):3873–3881.
Zhao D, et al. 2019. Mitochondrial haplogroups N9 and G are associated with metabolic syndrome among human immunodeficiency virus-infected patients in China. AIDS Res Hum Retroviruses. 35(6): 536–543.

Associate editor: Kirk Lohmueller