

ORIGINAL ARTICLE

The key role of patrilineal inheritance in shaping the genetic variation of Dagestan highlanders

Laura Caciagli¹, Kazima Bulayeva², Oleg Bulayev², Stefania Bertoncini¹, Luca Taglioli¹, Luca Pagani³, Giorgio Paoli¹ and Sergio Tofanelli¹

The Caucasus region is a complex cultural and ethnic mosaic, comprising populations that speak Caucasian, Indo-European and Altaic languages. Isolated mountain villages (*auls*) in Dagestan still preserve high level of genetic and cultural diversity and have patriarchal societies with a long history of isolation. The aim of this study was to understand the genetic history of five Dagestan highland *auls* with distinct ethnic affiliation (Avars, Chechens-Akkins, Kubachians, Laks, Tabasarans) using markers on the male-specific region of the Y chromosome. The groups analyzed here are all Muslims but speak different languages all belonging to the Nakh-Dagestanian linguistic family. The results show that the Dagestan ethnic groups share a common Y-genetic background, with deep-rooted genealogies and rare alleles, dating back to an early phase in the post-glacial recolonization of Europe. Geography and stochastic factors, such as founder effect and long-term genetic drift, driven by the rigid structuring of societies in groups of patrilineal descent, most likely acted as mutually reinforcing key factors in determining the high degree of Y-genetic divergence among these ethnic groups.

Journal of Human Genetics (2009) 54, 689–694; doi:10.1038/jhg.2009.94; published online 13 November 2009

Keywords: Caucasus region; Dagestan; genetic variation; patrilineal inheritance; Y chromosome; Y-SNP; Y-STR

INTRODUCTION

The Caucasus region provides a great opportunity to investigate the influence of geography and language on the genetic structure of human populations. The geomorphological and linguistic landscapes are extremely complex and heterogeneous with several ethnic groups speaking Caucasian, Indo-European or Altaic languages isolated over long periods in the region. There is no traditional pan-Caucasian lingua franca, and bilingualism or multilingualism are the main means of inter-ethnic communication.¹

Twenty-six out of the fifty autochthonous ethnic groups of the Caucasus live in Dagestan,² the north-eastern Caucasian Republic within the Russian Federation. Archeological data indicate that the earliest settlers in the Dagestan highland region are at least 10 000 years old.³ These ethnic groups can be considered as primary isolates,⁴ because they have been isolated and endogamous for centuries. Throughout the history of the region there has been restricted gene flow between neighboring populations, a high degree of inbreeding (reinforced by Muslim tradition of marriages within kindred), and relatively stationary population sizes for many generations.⁵ Prolonged marriage isolation under strict patrilocality favored the creation of genetically homogeneous gene pools within these highland populations, as well as a considerable genetic differentiation between them.⁶ Many groups show high prevalence of certain complex diseases such as

cardiovascular illnesses, cancer, schizophrenia, mental retardation and progressive muscular dystrophy.^{7,8}

Previous genetic studies have revealed that the Caucasian populations do not share the genetic variation believed by some researchers to be a signature of the Neolithic expansion through Europe, which has led some to infer that these populations are direct descendants of a more ancient Eurasian population.^{9–11} Others suggest that the genetic legacy of the Caucasus region is mainly because of the people who crossed or invaded this territory in historic times.¹²

A large amount of genetic data on the Caucasian populations, based on classical markers,^{9,13–15} *Alu* insertion polymorphisms,¹⁰ mitochondrial DNA^{10,16} and the non-recombining portion of the Y chromosome (NRY)¹⁷ have shown that neither geography nor linguistics have had a strong influence on the genetic structure of these ethnic groups and that genetic drift appears to be the main factor acting in the region. Furthermore, Y chromosomes of highland Dagestan ethnic groups appear to have undergone genetic drift independently of the population history of mitochondrial and autosomal loci, suggesting that social rules, and not simply geographic isolation, are the causal agents for the high NRY diversity observed among highlanders. However, these analyses are often based either on an insufficient number of ethnic groups or they lack sampling accuracy, because they do not consider only mono-ethnic settlements, nor do they use an

¹Dipartimento di Biologia, Università di Pisa, Pisa, Italy; ²Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia and ³Scuola Normale Superiore di Pisa, Pisa, Italy

Correspondence: Dr L. Caciagli, Dipartimento di Biologia, Università di Pisa, via Derna 1, Pisa 56100, Italy.

E-mail: Laura.Caciagli@gmail.com

Received 6 April 2009; revised 11 August 2009; accepted 1 September 2009; published online 13 November 2009

adequate number of molecular markers to convey a more accurate profile of the variation.

In this study, we analyzed both stable and fast-mutating polymorphisms on the male-specific region of the Y chromosome that gave a substantial increment in terms of evolutionary resolution with respect to previous analyses^{12,17,18} and provided a finer framework upon which evolutionary hypotheses were tested. Our main goal was to analyze the Y-chromosomal variation in these ethnic groups. In this way, it could be possible to better define the genetic structure of Dagestan ethnic groups, verifying whether the extreme genetic diversity of these populations is mainly due to stochastic factors (especially drift) or because of cultural and geographic barriers.

MATERIALS AND METHODS

Subjects and populations

Blood samples were collected in the summer of 2004 from 105 healthy, unrelated males living in the Dagestan *auls*. The sample set analyzed included five different ethnic groups: Tabasarans ($N=30$), Kubachians ($N=14$), Laks ($N=21$), Avars ($N=20$) and Chechens-Akkins ($N=20$) (Figure 1).

Appropriate informed consent was obtained from all individuals participating in the study. The project and subjects consent forms were approved by the Dagestan IRB at the Institute of History, Archaeology and Ethnology, Dagestan Center of the Russian Academy of Sciences. Sampling accuracy was ensured by choosing only mono-ethnic settlements, preferring isolated highland communities with social rigid structures influencing marriage behaviors and selecting males who were unrelated for at least three generations. The small sample size reflects the difficulty in obtaining an elevated number of unrelated samples in strictly endogamous, small-sized, isolated communities.

All sampled individuals were Muslims residing in remote villages located between altitudes of approximately 550 m (Chechens) and 2 118 m (Laks and Avars) above sea level. These ethnic groups are thought to be indigenous to the region but, like other native peoples of the Caucasus region, their exact origins are unclear. The Laks, Chechens, Avars and Tabasarans are farmers, transhumant shepherds and livestock breeders, whereas Kubachians are well-known goldsmiths and silversmiths. Social and family relations are regulated by *Adat* (traditional Dagestan customary law that pre-dates Islam) and *Shariah* (Islamic law). Except for the Chechens, who generally marry outside the clan, most marriages are between first and second cousins and take place within the *tukhum*, extended family units, or clans, composed of several families descending from a common male ancestor. This marriage practice controls the inheritance of property and restricts male gene flow.

Y-chromosome genotyping

Genomic DNA was extracted from whole blood using QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany), and genotyped for 20 Y-STR loci (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 a/b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, GATAC4, GATAH4 and YCAII a/b) and 30 Y-biallelic markers (Y *Alu* polymorphism, M96, M35, M78,

M89, M201, M285, M287, P15, M170, M253, P37.2, M223, M267, M62, M365, M367, M368, M369, M390, M172, M9, M22, M45, M17, P25, M124, M2, M90, MEH2) that were previously reported to be polymorphic in Europe and the Near East.^{19,20}

Seventeen Y-STRs were typed using the AmpFSTR Y-filer kit (Applied Biosystems, Foster City, CA, USA). The length of PCR products was estimated on the ABI-PRISM 310 Genetic Analyzer using the Genotyper version 3.7 analysis software (Applied Biosystems) and the Y-filer reference ladders. The internal size standard used was GeneScan 500 LIZ (Applied Biosystems). DYS388 and YCAII a/b loci were amplified separately. Primers labeled with different fluorescent dye labels, FAM (blue) and JOE (green), allowed the detection of these loci within the same allelic size range (between 125 and 143 bp for DYS388 and 135–163 bp for YCAII). Fragments were separated on the ABI 310 Genetic Analyzer using GeneScan 500 ROX size standard (Applied Biosystems).

Biallelic Y-markers were typed following a phylogenetic hierarchical approach to define evolutionary Y lineages, or haplogroups (Hg). The Y-Hg nomenclature used here is in accordance with the recommendations of the Y Chromosome Consortium.²¹ M223-I2b1 and M17-R1a1a subclades designation was updated to the Y-haplogroup classification of International Society of Genetic Genealogy (ISOGG 2009, version 4.13).

Genomic DNA was amplified for Y-SNPs using the ABI-PRISM SNaPshot Multiplex Kit. Primers were drawn from the literature and modified to construct multiplexes with the software Autodimer version 1. Each primer was checked for possible extendable hairpin structures and dimer formation between primers. PCR products were purified using Shrimp Alkaline Phosphatase and Exonuclease I to remove unincorporated dNTPs and primers. The purified fragments were separated and detected on ABI 310 Genetic Analyzer's capillary electrophoresis platform using the GeneScan 120 LIZ internal size standard.

Initially the Y *Alu* polymorphism was typed in all individuals and Y-SNPs M96, M35 and M78 were only typed in individuals with the insertion. Markers M170, M172, M267, M201 were typed only in individuals with the derived M89 allele and the ancestral state at M9 and M45; M22 was typed only in individuals with the mutant state at M89 and M9 and the ancestral state at M45, whereas M17, P25, M124 were typed in individuals with the derived allele at M89, M9 and M45. M62, M365, M367, M368, M369, M390 were typed only in individuals with the derived allele at M267; M285, M287 and P15 were typed only in individuals with the derived state at M201 and M253; P37.2, M223 were only typed in individuals with the mutant state at M170. M2, M90, MEH2, M253, P37.2, M62, M365, M369, M390, M285, M287 and P15 were genotyped, but not observed in any of the analyzed samples.

Statistical analysis

Analysis of molecular variance (AMOVA), genetic distances and diversity indices were calculated using Arlequin 3.11²² (<http://cmpg.unibe.ch/software/arlequin3>). The statistical significance of F_{st} values was estimated by permutation analysis, using 10 000 permutations. F_{st} values were calculated based on the number of pairwise differences between haplotypes/haplogroups. The statistical significance of the correlation between geographic distance, language and genetic distance matrices, was evaluated by AMOVA with 10 000 permutations. The relationship between populations was also assessed through a multidimensional scaling (MDS) analysis, based on haplogroup frequencies and performed with the STATISTICA version 6.0 (StatSoft, Inc., Tulsa, OK, USA).

A median-joining network²³ linking 20-locus haplotypes was constructed using the program Network 4.5.1.0 (<http://www.fluxus-engineering.com>). When constructing networks, the maximum value of 99 was assigned to each SNP whereas each STR locus was given a specific weight according to its variance; the weight of the i th STR was calculated as $10 V_m/V_i$, where V_m is the mean variance of all STRs and V_i is the variance of the i th STR.²⁴ DYS385 and YCAII were included in the networks as two separate loci.²⁴

Bayesian coalescent analysis was performed using the program BATWING, which uses a Markov-chain Monte Carlo method²⁵ to derive posterior distributions for a complete set of parameters that describe the relevant underlying model. The age of the most recent common ancestor of the lineage J1-M267 was calculated using this method as described earlier.¹¹



Figure 1 Geographic location of the Dagestan ethnic groups sampled.

We searched for rare or private alleles to demonstrate whether an ancient genetic structure is still preserved for the ethnic groups of Dagestan, comparing alleles experimentally found in the sample with those obtained from three databases:

- YHRD database (release 27, built on 13 February 2009, consisting of 72 082 haplotypes in 538 populations), (www.yhrd.org).
- Y-Chromosome database of Sorenson Molecular Genealogy Foundation (more than 31 000 haplotypes; last update at 26 February 2009), (www.smgf.org).
- Internal database made up by published and unpublished data (51 149 individuals, 31 453 of which are Europeans).

RESULTS

The main haplogroups detected in the five ethnic groups analyzed are shown in Figure 2, and allele frequencies are summarized in Supplementary Table 1s. Ten paternal lineages were observed: E1b1b1-M35, E1b1b1a-M78, G-M201, I-M170, J1-M267, J1e1-M367/M368, J2-M172, L-M22, R1a1a-M17, R1b1-P25. The most widespread haplogroup in the sample is haplogroup J1 (68.6%), in particular, J1x(J1a, J1b, J1c, J1e1, J1e2) and J1e1 (65 and 4%, respectively) followed by J2 (9.5%), R1b1 (5%), G, I and R1a1a (4%). No downstream mutations were observed in the majority of J1-M267 Y chromosomes and in the G-M201, I-M170 and E1b1b1-M35 lineages in these populations (Figure 3). This, and the presence of several rare or private alleles (Supplementary Table 2s) indicate the great antiquity of Y-chromosomal lineages in Dagestan ethnic groups and the long-term isolation of *auls*.

Nei's diversity estimates based on haplogroup frequencies, reveal low haplogroup diversity, with only Laks and Avars presenting a slightly higher diversity values (Supplementary Table 3s). In contrast, all Dagestan populations exhibited high haplotype diversity (Supplementary Table 3s), as expected in populations that have evolved under constant-size over several hundred generations.²⁴ In the Kubachians, which are one of the most strictly endogamous ethnic groups, a significant loss of paternal lineages was observed (haplogroup diversity=0.264 ± 0.136) and only two haplogroups, J1 and R1a1a, were found. However, they had the highest level of weighted mean intralinear mean pairwise difference (4.935) and haplotype diversity (HD=0.989 ± 0.031), with no haplotype sharing across 20 Y-STR loci.

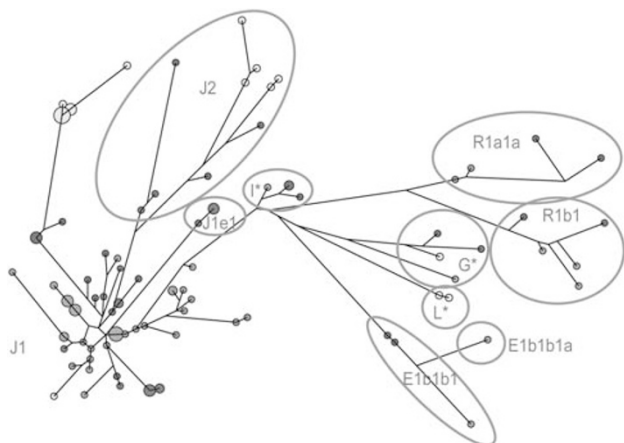


Figure 2 Median-joining network constructed using 20 locus STR haplotypes. Circles represent haplotypes, with areas proportional to the number of individuals they contain. Pink: Kubachians; red: Avars; yellow: Chechens; green: Tabasarans; light blue: Laks. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Considering the DYS388 and YCAII loci in the majority of individuals belonging to haplogroup J1, we did not detect the high repeat numbers (≥ 15) for DYS388 as reported in the Arab and Jewish populations from Near East nor did we find the single-banded motif YCAIIa22-YCAIIb22, associated with the diffusion of Arab people in Near East and North Africa.²⁶ The Cohen, Galilee and I&P Arab modal haplotypes,²⁷ which were reported in the Turkic populations from Anatolia¹⁹ were not found in the Dagestan samples.

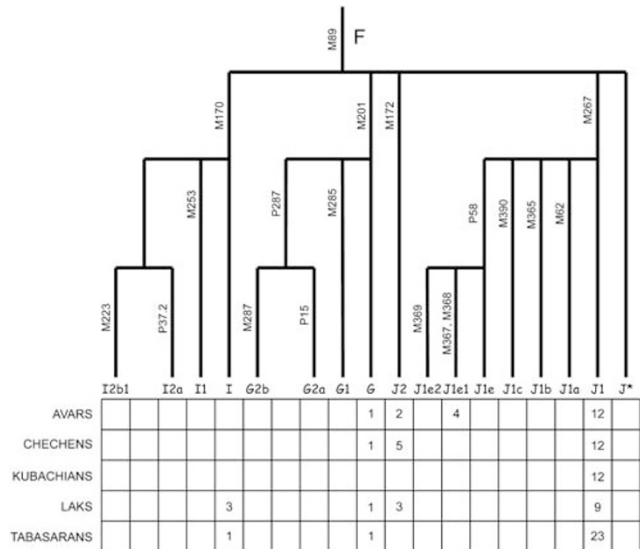


Figure 3 Y-chromosomal haplogroup frequencies in Dagestan ethnic groups, considering only the Y-Hg J, G, I and their subclasses.

Table 1 Reference populations along with their geographic locations, the number of individuals analyzed (N), linguistic affiliations and data sources

Geographic location	Populations	n	Linguistic family	References	
North-Caucasus	Avars	20	Nakh-Dagestani	This study	
	Laks	21	Nakh-Dagestani		
	Chechens	20	Nakh-Dagestani		
	Kubachians	14	Nakh-Dagestani		
	Tabasarans	30	Nakh-Dagestani		
	Abazinians	14	Northwest-Caucasian		12,18
	Non-Dag. Chechens	19	Nakh-Dagestani		
	Darginians(Dagestan)	26	Nakh-Dagestani		
	Ingushians	22	Nakh-Dagestani		
	Kabardinians	25	Northwest-Caucasian		
South-Caucasus	Lezginians(Dagestan)	25	Nakh-Dagestani		
	Ossetians(Ardon)	28	Indo-European		
	Ossetians(Digora)	31	Indo-European		
	Rutulians(Dagestan)	24	Nakh-Dagestani		
	Abkhazians	12	Northwest-Caucasian		
	Armenians	100	Indo-European		
	Azerbaijanians	72	Altaic		
Western Asia	Georgians	77	Kartvelian		
	Turks	39	Altaic		
	Iran (Tehran)	80	Indo-European		
	Iran (Isfahan)	50	Indo-European		
Total		783			

The ethnic groups analyzed in this study are in boldtype.

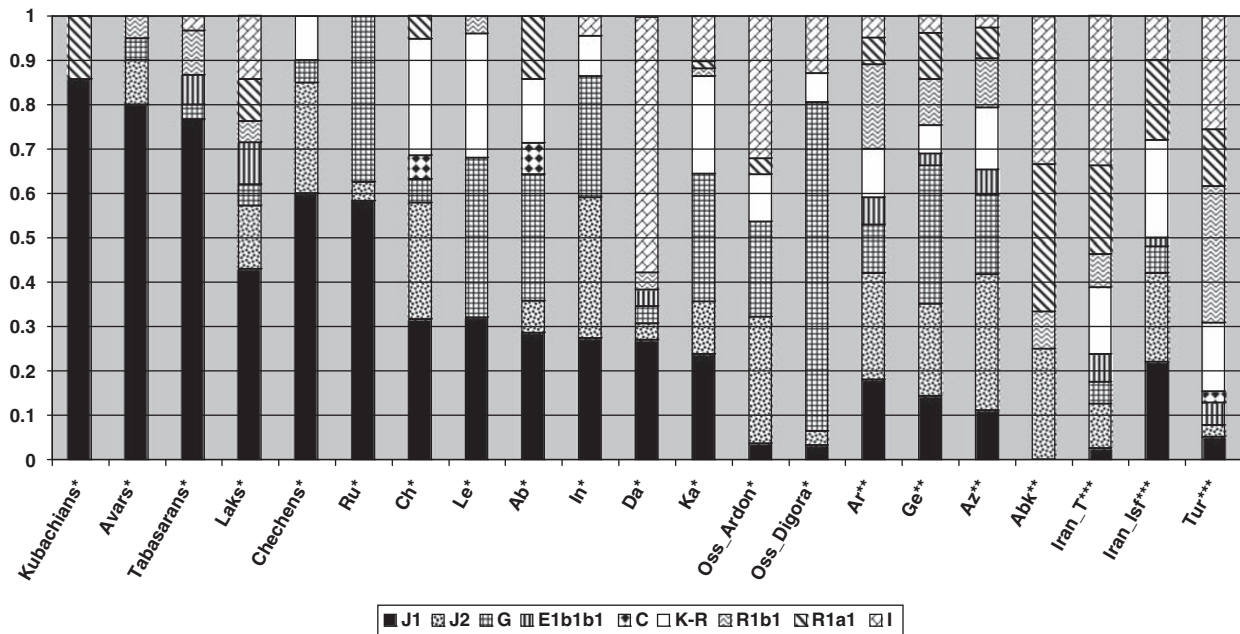


Figure 4 Y-Hg distribution in the samples analyzed. Abbreviations: Ab, Abazians; Abk, Abkhazians; Ar, Armenians; Az, Azerbaijanians; Ch, non-Dagestania Chechens; Da, Darginians; Ge, Georgians; In, Ingushians; Ka, Kabardinians; Le, Lezginians; Oss_Ardon, Ossetians from Ardon; Oss_Digora, Ossetians from Digora; Ru, Rutulians; Tur, Turks; Iran_Isf, Iranians from Isfahan; Iran_T, Iranians from Tehran. *North Caucasus; **South Caucasus; ***West Asia.

Genetic variability of the ethnic groups analyzed was further evaluated in the wider context of the Caucasus region. Populations were compared on the basis of haplogroup frequencies (Table 1). The most frequent and widespread haplogroup in north Caucasian populations was J1 (Figure 4). Ethnic groups from Dagestan have the highest frequencies ever reported for this haplogroup (mean value 69%) that range from 86% in Kubachians to 43% in Laks. J1 frequency in the non-Dagestania ethnic groups are significantly lower (mean value 20%), ranging from 32% (non-Dagestania Chechens) to 3% (Ossetians). This haplogroup is scarcely found in the South Caucasus (11%) where haplogroup J2 is more common (25%). In the samples analyzed here the J2 lineage was present only in Chechens (25%), Laks (14.3%) and Avars (10%), with a mean value not exceeding 10%. In non-Dagestania populations from North Caucasus this haplogroup is slightly more frequent (13%).

To visualize the relationships among Caucasians and West-Asian populations, we performed MDS analyses based on haplogroup frequencies (Figure 5). The Dagestan ethnic groups analyzed here are well separated from the remaining Caucasian and West-Asian populations in the first two dimensions. The North Caucasian ethnic groups analyzed in this study (Kubachians, Tabasarans and Avars) form a cluster that is distinct from the other northern (Kabardinians, Abazians, Ingushians and non-Dagestania Chechens) and southern (Georgians, Azerbaijanians, Armenians) Caucasian populations whereas the Chechens and Laks plot separately. Except for the outlier populations of Darginians and Ossetians (North Caucasus), all the other ethnic groups cluster together with populations of the same geographical area. Overall, there is a well-defined clustering by geography but no clear-cut population groupings by linguistic affiliation.

AMOVA was used to evaluate the level of genetic differentiation between populations (Supplementary Table 4s). Although the majority of haplotype variation was accounted for by within population differences (86.88%), the variance among populations component (13.12%) was highly significant ($P=0.00000 \pm 0.00000$). To examine

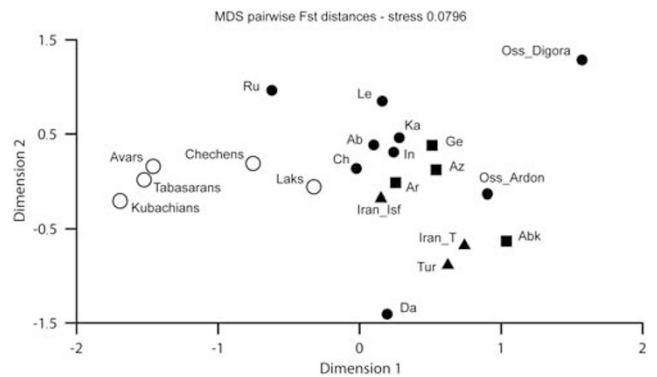


Figure 5 MDS plot based on the matrix of haplogroup frequencies, showing genetic relationships among Caucasus and West-Asian populations. Populations analyzed in this study are indicated by white circles and population designations are as described in Figure 4. Symbols used to designate geographic clusters: ■ South Caucasus; ● North Caucasus ▲ West Asia. The stress value for the MDS plot is 0.0796.

the genetic structure in Dagestan and surrounding populations, AMOVA was also performed on the paternal lineage frequencies using different grouping criteria (Supplementary Table 5s). Even if the more relevant percentage of variation was within populations (83.07 and 77.68%, respectively; $P=0.0000$), the variance between groups component was high and significant (9.75%; 14.94%; $P=0.0000$), underlining the peculiar characteristics of Dagestan ethnic groups.

The concordance between geography, language and genetic structure was further investigated by grouping populations according to geographic and linguistic criteria. The results (Table 2) show that a geographical rather than linguistic classification of Caucasian populations best accounts for the extant Y-chromosomal diversity. The within-groups component of the genetic variance was high and

Table 2 Analysis of molecular variance (AMOVA) according to geographic (RG) and linguistic (L) classification

Classifications	Percentages of genetic variance and fixation indexes					
	Among groups	Among populations within groups	Within populations	F_{CT}	F_{SC}	F_{ST}
RG	4.10	10.14	85.76	0.04105*	0.10570*	0.14241*
L	1.23	12.27	86.50	0.01226	0.12422*	0.13496*

RG=North Caucasus (14 populations: Tabasarans, Avars, Chechens, Laks, Kubachians, Abazinians, non-Dagestanian Chechens, Darginians from Dagestan, Ingushians, Kabardinians, Lezginians from Dagestan, Ossetians from Ardon and Digora, Rutulians from Dagestan); South Caucasus (four populations: Abkhazians, Armenians, Azerbaijanians, Georgians); West Asia (Turks, Iranians from Tehran and Isfahan).

L=Caucasian languages (Abazinians, Abkhazians, non-Dagestanian Chechens, Darginians from Dagestan, Georgians, Kabardinians, Ingushians, Lezginians from Dagestan, Rutulians from Dagestan, Tabasarans, Laks, Kubachians, Avars, Chechens); Indo-European languages (Armenians, Ossetians from Ardon and Digora, Iranians from Tehran and Isfahan); Turkic languages (Azerbaijanians, Turks). * $P < 0.0001$, significance tests: 10 000 permutations.

statistically significant: for geographically defined groups (RG: North Caucasus, South Caucasus, West Asia), F_{SC} was 0.10570 ($P = 0.00000 \pm 0.00000$); for linguistically defined groups (L: Caucasian, Indo-European, Turkic), F_{SC} value was 0.12422 ($P = 0.00000 \pm 0.00000$). In contrast, the among-groups component (F_{CT} value) was high and significantly different from zero only for geographical classification ($F_{CT} = 0.04105$, $P = 0.02673 \pm 0.00153$) and not for the linguistically defined group ($F_{CT} = 0.01226$; $P = 0.16109 \pm 0.00379$).

DISCUSSION

The Republic of Dagestan (Northern Caucasus, Russia) has a mosaic of ethnic groups with claims to long residence in the area.²⁸ The linguistic and ethnic Caucasian diversity is particularly apparent in the small villages, or *auls*, along the often inaccessible highland region of the Great Caucasus Range.

This study presents the most comprehensive analyses of NRY markers in the Dagestan highland ethnic groups. The sampling strategy of this study was designed to obtain a realistic portrayal of the genetic structure and variability of all the ethnic groups analyzed, even if the sample size was low. We analyzed both stable and fast-mutating polymorphisms, which respond differently to demographic events and time scales and allowed a finer resolution of Y variation in the region. The results show that the patrilineal descent of the Dagestan groups represents a *cul-de-sac* in which NRY lineages (haplogroups/haplotypes) that originated during the Last Glacial Maximum (LGM) are still preserved. These *auls* may have functioned as refuges off the expansion routes of human populations since the early Neolithic age.

Overall, the Dagestan groups analyzed here represent outliers in the Caucasus Y genetic landscape because of the presence of rare alleles and prevalence of M267-derived J1 Y chromosomes. The absence of downstream mutations in clades J1-M267 (with the only exception of the J1e1 subclade), G-M201, I-M170 and E1b1b1-M35 chromosomes (Figure 3), may represent both the great antiquity of Y-paternal lineages of these ethnic groups and genetic drift in an isolated population. Haplogroup frequencies reported here are very different from those of previous studies¹² (Figure 4). North, South Caucasus and West-Asian populations analyzed showed high frequencies for haplogroups J2-M172, G-M201 and I-M170, whereas in Dagestan these haplogroups are found in <10% of the population.

The genetic heterogeneity previously found in Caucasian populations did not correlate with either linguistic diversity or geographic barriers in the Caucasus.¹² Based upon mitochondrial DNA and Y-markers the Indo-European speaking Armenians and Turkic-speaking Azerbaijanians were genetically more closely related to other Caucasus groups and not to other Indo-European or Turkic-speaking

groups.^{12,29} Although geographic classifications do provide a better fit for the genetic variability, correlations between genetic and geographic distances between groups were not statistically significant.¹² Our results depict a significantly different scenario with the Dagestan ethnic groups showing a unique high genetic diversity derived from an ancient founder pool of haplotypes. Linguistic barriers do not appear to be the main determinants of the high between-population Y-chromosome variability observed in the Caucasus (Table 2, Figure 5). Instead, geographic factors appear to have shaped the high genetic diversity observed in this region. Ethnic groups from Northern Caucasus are often small-sized, isolated populations residing in highland regions and, hence, they are likely to have undergone long-term genetic drift. In the populations from Southern Caucasus gene flow, facilitated by the absence of strong geographic barriers, might have reduced the genetic distance between groups.

Time estimates were obtained¹¹ with BATWING, considering a constant-size population growth model. The time to the most recent common ancestor of a set of closely related Dagestan Y chromosomes belonging to haplogroup J1 was estimated to be 11 711 years (95% CI: 15 460–8930, using as priors a uniform distribution between germline and evolutionary estimated mutation rates and 25 years as inter-generation time). This is consistent with the events of spatial re-expansion that took place after the last glacial maximum in the northern hemisphere, the latest being caused by the end of the Younger Dryas cooling (12 900–11 600 ybp³⁰). Moreover, this estimate is in agreement with the archeological record which indicates that some of these ethnic groups have lived in the Dagestan mountains for over 10 000 years.^{2,31} These results suggest that colonization of Dagestan probably took place during an early phase of the peopling of Eurasia.

With respect to Semino's and Nebel's hypotheses^{26,27} it is likely that the Islamic diffusion in historical time involved the Dagestan population but only from a cultural point of view. Numerous Near Eastern groups have invaded the Caucasus several times during the last two millennia, including the Arab caliph's occupation of Georgia in 654 AD, the Seljuk Turks invasion of the South Caucasus in the 11th century and repeated invasions by Turks and Persians.^{20,32} Nevertheless, our results show that these did not result in any genetic imprint probably because the Great Caucasus Range might have represented an important genetic barrier to recent migrations. Geographic and cultural (the rigid structuring of group of patrilineal descent, the high rate of endogamy and inbreeding, the small population size of some ethnic units) factors, could not prevent the diffusion of Islam in 12th–14th AD, but may have represented a genetic watershed in historical times.

In conclusion, Dagestan highland villagers represent a unique example of biodiversity resulting from over four hundred generations

of independent evolution. Presumably, all share a common Y genetic background dating back to an early phase of the post-glacial recolonization of Europe. Subsequent geographic isolation³³ and endogamy of highland populations resulted in their high genetic differentiation. These factors have had a key role in shaping the present gene pools in Dagestan by facilitating the emergence of a high genetic divergence among ethnic groups and the maintenance of both rare allelic variants and ancestral Y-chromosomal lineages.

ACKNOWLEDGEMENTS

This work was supported by a MIUR COFIN Grant, N.2003054059-002 to GP. We thank Dr Silvia Agostini for useful discussions and suggestions and are grateful to Dr Vittoria Carnicelli for her technical assistance. Our most sincere gratitude, however, must be reserved for Dr Qasim Ayub of The Wellcome Trust Sanger Institute, who has reviewed the entire paper and offered constructive and helpful criticism.

- 1 Nichols, J. *The Nakh Dagestani consonant correspondences* 207–251 (John Benjamins, Philadelphia, 2003).
- 2 Bulayeva, K., Jorde, L., Hostler, C., Bulayev, O. A., Pavlova, T. A. & Harpending, H. STR polymorphism in indigenous Dagestani populations. *Russian J Genet* **40**, 558–569 (2004).
- 3 Gadjević, M., Davudov, O. M. & Shikhsaidov, S. History of the Dagestan. *Nauka (Science)* **45**, 345 (1996).
- 4 Neel, J. V. *Minority populations as genetic isolates: the interpretation of inbreeding results*. 1–13 (Macmillan, London, 1992).
- 5 Bulayeva, K. Genetic basis of human psychophysiology. *Nauka (Science)* **95**, 218–223 (1991).
- 6 Bulayeva, K. B., Jorde, L., Watkins, S., Ostler, C., Pavlova, T. A., Bulayev, O. A. *et al*. Ethnogenomic diversity of Caucasus, Dagestan. *Am J Hum Biol* **18**, 610–620 (2006).
- 7 Bulayeva, K. B. Overview of genetic-epidemiological studies in ethnically and demographically diverse isolates of Dagestan, Northern Caucasus, Russia. *Croat Med J* **47**, 641–648 (2006).
- 8 Bulayeva, K. B., Leal, S. M., Pavlova, T. A., Kurbanov, R. M., Glatt, S. J., Bulayev, O. A. *et al*. Mapping genes of complex psychiatric diseases in Dagestani genetic isolates. *Am J Med Genet B Neuropsychiatr Genet* **132B**, 76–84 (2005).
- 9 Barbuĵani, G., Nasidze, I. S. & Whitehead, G. N. Genetic diversity in the Caucasus. *Hum Biol* **66**, 639–668 (1994).
- 10 Bulayeva, K., Jorde, L. B., Ostler, C., Watkins, S., Bulayev, O. & Harpending, H. Genetics and population history of Caucasus populations. *Hum Biol* **75**, 837–853 (2003).
- 11 Tofaneli, S., Ferri, G., Bulayeva, K., Caciagli, L., Onofri, V., Taglioli, L. *et al*. J1-M267 Y lineage marks climate-driven pre-historical human displacements. *Eur J Hum Genet* advance online publication, 15 April 2009; **58**, 1–5 (2009).
- 12 Nasidze, I., Ling, E. Y., Quinque, D., Dupanloup, I., Cordaux, R., Rychkov, S. *et al*. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* **68**, 205–221 (2004).
- 13 Barbuĵani, G., Whitehead, G. N., Bertorelle, G. & Nasidze, I. S. Testing hypotheses on processes of genetic and linguistic change in the Caucasus. *Hum Biol* **66**, 843–864 (1994).
- 14 Bulayeva, K., Dubinin, N. P., Shamov, I. A., Isaichev, S. A. & Pavlova, T. A. Populatsionnaia genetika gortsev Dagestana. *Genetika* **21**, 1749–1758 (1985).
- 15 Dubinin, N. P. & Bulaeva, K. B. Genetic bases of individuality in human populations. *Dokl Akad Nauk SSSR* **265**, 470–473 (1982).
- 16 Nasidze, I. & Stoneking, M. Mitochondrial DNA variation and language replacements in the Caucasus. *Proc Biol Sci* **268**, 1197–1206 (2001).
- 17 Nasidze, I., Sarkisian, T., Kerimov, A. & Stoneking, M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet* **112**, 255–261 (2003).
- 18 Nasidze, I., Schädlich, H. & Stoneking, M. Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Sci Int* **137**, 85–93 (2003).
- 19 Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L. *et al*. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* **114**, 127–148 (2004).
- 20 Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E. *et al*. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159 (2000).
- 21 Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* **18**, 830–838 (2008).
- 22 Excoffier, L., Laval, G. & Schneider, S. Arlequin (ver. 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50 (2005).
- 23 Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48 (1999).
- 24 Bosch, E., Calafell, F., Gonzalez-Neira, A., Flaiz, C., Mateu, E., Scheil, H. G. *et al*. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* **70**, 459–487 (2006).
- 25 Wilson, I. J. & Balding, D. J. Genealogical inference from microsatellite data. *Genetics* **150**, 499–510 (1998).
- 26 Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V. *et al*. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**, 1023–1034 (2004).
- 27 Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M. & Oppenheim, A. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* **69**, 1095–1112 (2001).
- 28 Gadjević, A. Anthropology of small Dagestani populations. *Dagestan Branch of USSR Academy of Sciences, Makhachkala* **25**, 216–221 (1971).
- 29 Nasidze, I., Risch, G. M., Robichaux, M., Sherry, S. T., Batzer, M. A. & Stoneking, M. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* **9**, 267–272 (2001).
- 30 Firestone, R. B., West, A., Kennett, J. P., Becker, L., Bunch, T. E., Revay, Z. S. *et al*. Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. *Proc Natl Acad Sci USA* **104**, 16016–16021 (2007).
- 31 Gadzhiev, M. G., Davudov, O. M. & Shikhsaidov, S. M. Istorija Dagestana. *Nauka (Science)* **33**, 115–120 (1996).
- 32 Muskhelishvili, D. L. A historical atlas of Georgia. *Bull Georg Natl Acad Sci* **175**, 156–160 (2007).
- 33 Destro Bisol, G., Anagnostou, P., Batini, C., Battaglia, C., Bertoni, S., Boattini, A. *et al*. Italian isolates today: geographic and linguistic factors shaping human biodiversity. *J Anthropol Sci* **86**, 179–188 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)