

## Phylogenetic Relations of the Various G sub-haplogroups in Azeris and the Development of Early Agriculture in the Northwest of Iran

Ardeshir Bahmanimehr

Institute of Molecular Biology, National Academy of Science of Armenia

**Abstract:** Based on contemporary DNA distributions and cautious inferences on migratory histories, considering previously localized the potential geographic origin of haplogroup G-M201, by using high-resolution phylogenetic relationships containing both G1-M285- and G2-P287- related lineages, this research tries to find some genetic evidences to trace the spread of early Neolithic agriculture in the Northwest of Iran. For Iranian Azeris, the presence of haplogroup G-M201 has been associated with the spread of agriculture, as it reported especially for the European context (King RJ, et al., 2011). Considering general frequency pattern of haplogroup G in Iranians, it could be inferred that mountainous and unique harsh geography of west and northwest of Iran might facilitated the establishment of specific lineage of Neolithic agriculturalists, e.g. G haplogroup, in the region. However uninhabited Great Salt Desert created particular sharp border in the center and east of the country and cut further distribution of these lineages to the eastern neighbors and east part of Iranian plateau. Diversity of subhaplogroup G2-P287 among Iranians and coalescent time of the lineage formation in the region reflect a phylogenetic branching event that is in consistency with historical events. In this study, the distribution of informative G sub-clades from the Near/Middle East and the Caucasus has been studied.

[Ardeshir Bahmanimehr. **Phylogenetic Relations of the Various G sub-haplogroups in Azeris and the Development of Early Agriculture in the Northwest of Iran.** *Life Sci J* 2013;10(6s):517-521] (ISSN:1097-8135). <http://www.lifesciencesite.com>. 80

**Key words:** Haplogroup G, Iranian Azeris, Y chromosome,

### Introduction

Haplogroup G is believed to have originated around the Middle East during the late Paleolithic, possibly as early as 30,000 years ago (King RJ, et al., 2011). At that time, humans would all have been hunter-gatherers, and in most cases lived in small nomadic or semi-nomadic tribes. Members of this haplogroup appear to have been closely linked to the development of early agriculture in the Levant part of the Fertile Crescent, starting 11,500 years before present (Balanovsky O, et al., 2011).

It is thought that early Neolithic farmers have spread from the Levant westward to Anatolia and Europe, eastward to Mesopotamia and South Asia, and southward to the Arabian Peninsula and North and East Africa. The domestication of goats and cows first took place in the mountainous region of eastern Anatolia, including the Caucasus and Zagros Mountains. This is probably where the roots of haplogroup G2a (and perhaps of all haplogroup G) are to be found; however, there are several theories regarding the origin of G2a in Europe.

High level of genetic diversity detected in Iranian Azeris (Grugni V, et al., 2012) might be explained by the peculiarity of the region as a corridor for ancient human migrations and conducive conditions for sedentary habitation leading to sharp demographic growth in the area. Furthermore, environmental fluctuations occurring over time have changed areas feature and influence human life and

migrations, as a unique geography of Northwest of Iran that might facilitated the establishment of social organizations possibly caused population density in this area, extended to west foothills of Zagros Mountains. In whole, due to its pivotal geographical location and proximity to the Fertile Crescent the Northwest of Iran was a main migratory passage for the Neolithic agricultural diffusion into both western and eastern Eurasia.

Here, we are intending to trace back the itineraries of early Neolithic farmers in the Northwest of Iran through the spatial distribution of some major haplogroups of Y chromosome.

We admit that a comprehensive survey of the geographic distribution patterns of the haplogroup G sub-clades may provide useful information on different migratory events and demographic processes.

### Material and Methods

A total of 297 samples from deep area of three provinces of northwest Iran (East Azerbaijan, Ardebil and West Azerbaijan) were collected. Genomic DNA was extracted from whole blood by using standard procedure and protocol of Qiagene (QIAamp DNA Mini Kit) blood extraction kit. Samples were genotyped with a set of 58 Y-chromosomal binary markers on the non-recombining portion of the Y-chromosome that have been previously reported in ISOGG 2012 (<http://www.isogg.org/tree/>). All haplogroup G (M201-derived) samples were genotyped in a

hierarchical manner for the following binary markers: M285, P20, P287, P15, L91 P16, M286, P303, U1, L497, M406, Page19, M287 and M377, also five new markers: M426, M461, M485, M527 and M547 introduced in our previous publication (Rootsi S, et al., 2012). Furthermore, markers Page94, U5, U8 and L30 were typed in contextually appropriate samples to establish the position of the five new markers within the phylogeny. We genotyped binary markers following PCR amplification, by either Denaturing High Performance Liquid Chromatography, RFLP analysis, Taqman assay (Applied Biosystems, Foster City, CA, USA) or direct Sanger sequencing methodology.

All M201-derived samples were typed for short tandem repeats of Y-chromosome (Y-STRs) using 17 STR markers in the Applied Biosystems Amp FISTR Yfiler Kit according to the manufacturer recommendations. Two additional markers, DYS388 and DYS461, were typed separately. The fragments were run on the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). The results were analyzed using the ABI PRISM program Gene Mapper 4.0 (Applied Biosystems).

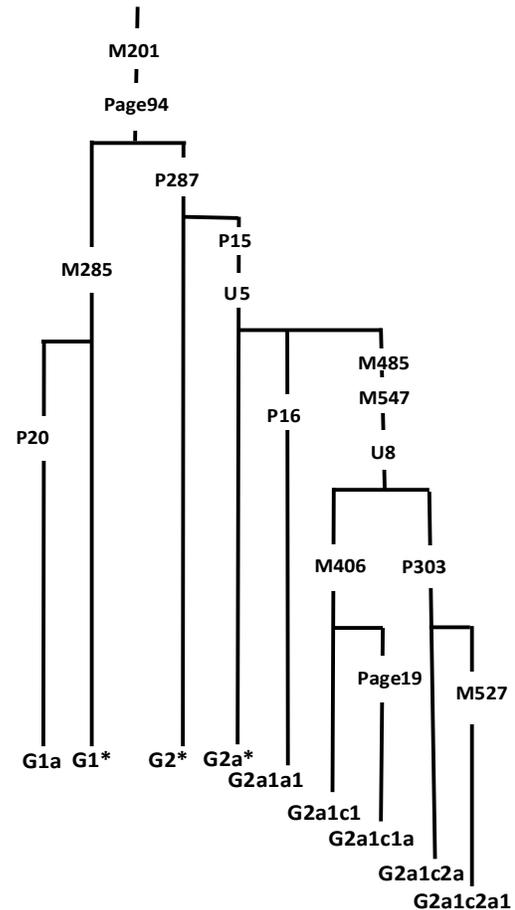
### Results

There are at least 42 distinct lineages, or subclades, reported within Haplogroup G (ISOGG 2012). These subclades are smaller lineages derived from two haplogroup G branches: G1 and G2. While using Y-STR data, the Td expansion time for all combined G1- M285 and G2- P287 affiliated chromosomes were estimated to be  $19,271 \pm 6,158$  and  $1,6356 \pm 2,264$  years ago respectively (Rootsi S, et al., 2012).

Haplogroup G is the fourth major haplogroup in Iranian Azeris with total frequency of 13%. As it is clear in the phylogenetic relations of the various sub-haplogroups investigated in the Figure.1, notably no basal G-M201\*, Page94\*(xM285, P287) chromosomes were detected in Iranian Azeris. In our samples just haplogroups G1\*(M285), G1a (P20), G2\*(P287), G2a\*(P15), G2a1a1 (P16), G2a1c2a1 (P303), G2a1c2a1 (M 527), G2a1c1 (M406), G2a1c1a (page 19) were detected.

The lineage G1 is the South and Central Asian branch of haplogroup G. G2a men migrated west to Anatolia and Europe in the Neolithic, while their G1 cousins migrated east to Persia and India (Balanovsky O, et al., 2011). Only very rare cases of G1 have been found in Europe, including Britain, Germany, as well as most of Southern, Central and Eastern Europe. Its frequency distribution extends further eastward as far as Central Asian Kazakhs. While the present-day frequency of G1 is low across its spread zone, in some Iranian ethnic groups as Iranian Mazandarani, Baluch and Kurds it is frequent

(Table.1) the expansion time estimate for this lineage attests to considerable antiquity.



**Figure 1.** Phylogenetic relations of studied binary markers within haplogroup G in Iranian Azeris.

In contrast to G1, the absolute majority of Iranian haplogroup G belongs to G2-P287 related subclades, with the vast majority of them being associated with G2a-P15-related lineages. The G2 clade consists of one widespread but relatively infrequent collection of P287\*, M377, M286 and M287 chromosomes versus a more abundant assemblage consisting of G2a-related P15\*, P16 and M485-related lineages.

G2a-P15\* was observed as the most common lineage among Iranian Azeris (4%). The expansion zone of this lineage was reported across a wide area extending from the Near/Middle East to the Balkans and Western Europe in the west, the Caucasus (especially the South Caucasus) in the north and Pakistan in the east (Yunusbaev B, et al., 2011). Haplogroup G men, who are negative for all G2a

subgroups, are uncommon in Europe. While this lineage is very frequent in the Zagros Mountains range, western (Arab, Lor, Kurd) and northwestern (Azeris, Gilan, Mazandaran) part of Iran. In contrast, in the Eastern population (Baluch) it is completely absent or it is appeared in the low frequency in the northeast of Iran (Turkmens). Frequent distribution of G2a-P15\* across west and northwest of Iran is in compliance with its expansion zone in the neighboring of south Caucasus. Although its downstream L91 sub-lineage is completely absent in Iranian ethnic groups.

The Caucasus specific G2a1-P16 lineage which effectively accounts for about one-third of the Caucasian male gene pool (Rootsi S, et al., 2012; Balanovsky O, et al., 2011; Cinnioglu C, et al., 2004; Flores C, et al., 2005) is essentially absent in Iranians with the exception of Gilan. The coalescence age estimate of 9,400 years for G2a1a-P16 coincides with the early Holocene.

The only ethnic group that has a majority of this haplogroup nowadays is the Ossetians in the North Caucasus, in the modern Russian Republic of North Ossetia-Alania. They are thought to descend directly from the Alans, a Central Asian tribe related to the ancient Scytho-Sarmatian who spoke an Eastern Iranian language. In some researches, European G2a1a individuals were considered as direct descendants of the ancient Alani, a tribe that invaded the Roman Empire in the late 4<sup>th</sup> and early 5<sup>th</sup> centuries (King RJ, et al., 2011).

G2a1c-M485 is the next most frequent and widespread lineage in Iranian Azeris. Among its several subgroups P303, M406 and M527 are observed almost in equal rate in Iranian Azeris.

The overall P303-derived chromosomes are the most widespread of all haplogroup G lineages. Compared with G1-M285, the phylogenetic level of P303 (Figure1) is shallower. In Iranian Azeris, P303 is the second frequent sublineage of haplogroup G. The highest frequency values for P303 are reported in the populations of the Caucasus region, being especially high among South Caucasian Abkhazians (24%) and Northwest (NW) Caucasian Adyghe and Cherkessians - 39.7% and 36.5%, respectively. In the Near/Middle East, the highest P303 frequency is detected among Palestinians (17.8%), whereas in Europe its frequency does not exceed 6%. Total collective coalescent age estimated for P303 is 12, 600 years (Rootsi S, et al., 2012).

Although geographic spread zone of G2a-P303 covers the whole haplogroup G distribution area, and contrarily to other branches of G2a, which are more prevalent in mountainous areas, G2a-P303 is found uniformly throughout Europe even in Scandinavia and Russia.

The coalescence age of G2a-P303 is estimated to be 6,700YBP for Iranians (Rootsi S, et al., 2012) which matches the time of the Indo-European expansion during the Bronze Age. More importantly, the presence of this lineage in India, especially among the upper castes, and combined presence across Europe is still a very strong argument in favor of an Indo-European origin for G2a-P303 (Kivisild T, et al., 2003; Yunusbaev B, et al., 2011).

Another frequent subclade of the G2a1c-M485 lineage is G2a1c1-M406. In contrast to its widely dispersed sister clade defined by P303, haplogroup G-M406 has a peak frequency in Cappadocia, Mediterranean Anatolia and Central Anatolia (6–7%) and it is not detected in most other regions with considerable P303 frequency (Ref.?). The expansion time of G-M406 in Anatolia is 12,800 years ago, which corresponds to climatic improvement in the beginning of the Holocene and the commencement of sedentary hunter-forager settlements at locations, such as Gobekli Tepi in Southeast Anatolia, thought to be critical for the domestication of crops (wheat and barley) that propelled the development of the Neolithic. More distantly, G2a3a-M406 occurs in Italy (3%) with a coalescent age (Td) of 8,100 years ago, consistent with the model of maritime Neolithic colonization of the Italian peninsula from coastal Anatolia and/or the Levant. Finally, to the east, G2a3a-M406 has an expansion time of 8,800 years ago in Iran (Rootsi S, et al., 2012), a time horizon that corresponds to the first Neolithic settlements in the Zagros Mountains of Iran. Thus, G2a3a-M406, along with other lineages such as J2a3b1-M92, may track the expansion of the Neolithic from Central/Mediterranean Anatolia to Greece/Italy and Iran.

Furthermore, the G2a1c-M485 and U1-specific subclade M527 are completely absent in Iranians and all neighboring region as Anatolia and the Caucasus while observed around 2% in Iranian Azeris. M527 is the most pronounced lineage among Ukrainians and Anatolian Greeks and it may reflect subsequent Greek maritime Iron Age colonization events (King RJ, et al. 2011) and perhaps, given its appearance among the Druze and Palestinians, even episodes associated with the enigmatic marauding Sea Peoples (Kaniewski D, et al., 2011).

This study took place in the mountainous region of eastern Anatolia, including the Caucasus and Zagros Mountains. This is probably where the roots of haplogroup G2a (and perhaps of all haplogroup G) are to be found. This would explain why haplogroup G is more common in mountainous areas, whether in Europe or in Asia, which is confirmed by the fact that Neolithic farmers in Europe belong primarily (even

exclusively if assimilated local population are omitted) to the haplogroup G2a (King RJ, et al. 2011).

**Table 1.** Haplogroup G and its sub-clades frequencies (%).

HG	G1	G1a	G2*	G2a*	G2a	G2a1	G2a3	G2a3a	G2a3b*	G2a3b2a	G2c	Pop Size	REF
<b>Defining Marker</b>	M285	P20	P287*	P15*	L91	P16	M406	PGE19	P303*	M527	M377		
<b>Iran</b>	1.8		0.9	3.5	0.1	0.2	0.7	0.1	3.3	0.2	0.1	938	5
<b>Ir-Gilan</b>	1.6			3.1		1.6			<b>9.4</b>			64	5
<b>Ir-Turkman</b>				1.4			1.4		2.9			68	5
<b>Ir-Persian</b>	1.2		0.62	3.1			2.5		3.75			160	5
<b>Ir-Arab</b>			1.8	10.5			1.8					57	5
<b>Ir-Kurd</b>	3.4		3.4	5.1								59	5
<b>Ir-Lor</b>	2			3.9					<b>9.8</b>			50	5
<b>Ir-Maz</b>	4.2			4.2					<b>11.1</b>			72	5
<b>Ir-Baluch</b>	4.2											24	5
<b>Ir-Azeri</b>	0.3	1	0.3	<b>4</b>		0.3	<b>2</b>	0.3	<b>2.4</b>	<b>2</b>		<b>297</b>	*
<b>Afghan</b>				1		0.5					<b>5.2</b>	190	6
<b>Qatar</b>				2.8					1.4			72	3
<b>UAE</b>	1.8	0.6		1.8			0.6					164	3
<b>Yemen</b>				1.6			1.6					62	3
<b>S-Caucasus</b>	2			41.3								305	14
<b>N-Ossetians</b>				1.5		63.6			2.3			132	14
<b>S-Ossetians</b>						43.5						23	14
<b>Turkey</b>	0.2	1	0.2	8.4		1						523	4
<b>Pakistan</b>	0.6			1.1			0.6		2.8		1.1	176	13
<b>Parsis (India)</b>									2.4			41	9
<b>Armenians</b>	1.6	0.9		1.4	0.9	1.2	2.1	0.2	1.6	0.2		426	12
<b>Askenazi Jews</b>	0.6					0.6					<b>5.1</b>	157	2

About all G2-P287 lineages in Afghanistan are G2b-M377. It has also been found in Ashkenazi Jews (Behar DM, et al., 2010), probably as an offshoot of Neolithic farmers from the Levant. In this study, it was found that overall haplogroup G2-P287 frequency drops abruptly from about 17.6% in Iranian Arabs, 13.7% in Loris and 11.3% in Azeris from West part of Iran, about 16.7% in Gilan and 14.1% in Mazandaran from the North part of the country to complete absence in Iranian Baluchs from eastern part and low frequency in the eastern neighbors, Afghanistan and

Pakistan (6.7% and 5.6%, respectively), marking the eastern border of the G2-P287 lineage's distribution. It seems that two uninhabited Great Salt Deserts in the middle of the Iranian plateau, stretching from the Alborz Mountain range in the northwest to the southeast of the country, plays a role of a particular sharp border which cut the distribution of these Neolithic lineages eastward. It also might emphasize the important role of fertile region in the north of Iran as a gateway for gene flow through/against the geographical barriers in the West and East of Iran.

## Conclusion

The topography of Iran consists of rugged, mountainous rims surrounding high interior basins. The mountains enclose several broad basins, or plateaus, on which major agricultural and urban settlements are located.

The geographic origin of haplogroup G has reported somewhere nearby eastern Anatolia, Armenia or western Iran (Rootsi S, et al., 2012). Also due to the general frequency pattern of haplogroup G, overall seems that the spread of this haplogroup extends over an area from southern Europe to the Near/Middle East and the Caucasus, but then decreases rapidly toward southern and Central Asia.

The majority of haplogroup G samples in the Iranian Azeris belongs to subhaplogroup G2 and reflects a branching event that is phylogenetically distributed among other Iranians. Overall contrast distribution of G2-P287 on both sides of Dasht e-Kavir and Dasht e-Lut deserts is an important evidence of a barrier role for these geographical boundaries. These deserts could be considered as a particular sharp border that cut the distribution of specific lineages of Neolithic farmers to the east part of Iranian plateau. It also might emphasize the important role of fertile region in the north of Iran as a gateway for gene flow through/against the geographical barriers in the West and East of Iran.

## References

- 1- Balanovsky O, Dibirova K, Dybo A et al. Parallel evolution of genes and languages in the Caucasus region. *Mol Biol Evol.* 2011; 28: 2905–2920
- 2- Behar DM, Yunusbayev B, Metspalu M et al. The genome-wide structure of the Jewish people. *Nature.* 2010; 466: 238–242.
- 3- Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ. Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Gen.* 2008; 16:374–386.
- 4- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet.* 2004; 114:127–148.
- 5- Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Semino O et al. Ancient Migratory Events in the Middle East: New Clues from the Y Chromosome Variation of Modern Iranians. *PLoS ONE.* 2012; 7(7): e41252. doi:10.1371/journal.pone.0041252.
- 6- Herrera KJ, Lowery R, Hadden L, Calderon S et al. Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists. *Eur J Hum Gen.* 2012; 20: 313–320
- 7- Kaniewski D, Van Campo E, Van Lerberghe K et al. The Sea Peoples, from cuneiform tablets to carbon dating. *PLoS One* 2011; 6: e20232.
- 8- King RJ, Di-Cristofaro J, Kouvasi A. The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. *BMC Evol Biol.* 2011; 11: 69.
- 9- Kivisild T, Rootsi S, Metspalu M et al. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet.* 2003; 72:313–332.
- 10- Nasidze I, Quinque D, Rahmani M, Alemohamad S, Stoneking M. Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. *Ann Hum Genet.* 2008; 72:241–252
- 11- Nei M. *Molecular Evolutionary Genetics.* Columbia University Press, New York, USA, 1987.
- 12- Rootsi S, Myres NM, Alin A, Jarve M, King R, Villems R et al. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur J Hum Gen.* 2012; 1–8.
- 13- Sengupta S, Zhivotovsky LA, King R et al. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet.* 2006. 78: 202–221.
- 14- Yunusbayev B, Metspalu M, Villems R et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol.* 2011; doi: 10.1093/molbev/msr221.

4/6/2013