DIAGNOSTIC Y-STR MARKERS IN HAPLOGROUP G

Phillip G. Goff and T. Whit Athey

Abstract

Y-Chromosome Haplogroup G reaches its highest frequency in the Caucasus Region (70% in N. Ossetia) and decreases in frequency in Western Europe to about one-to-two percent of the population on the Atlantic coast. Haplogroup G, like its brother haplogroups

genealogical research has steadily increased since 2000.

[1]

As of December 2005, about 40,000 genealogically-relevant haplotypes are available through various online databases.[2] Many who have been tested for their Y-STR haplotype want to know their predicted haplogroup with some level of certainty before taking a SNP-test. The identification of diagnostic Y-STR markers will help to fill this demand.

Members of Y Haplogroup G have repeat values on several Y-STR markers that are distinctively different from those of other haplogroups. These markers include DYS425, DYS452, DY S446, and DYF399S1. In

DYS425 is part of a larger marker called DYF371. DYF371 has four alleles, three of which have a C base in a particular location adjacent to the repeat structure, and fourth has a T base in that location. DYS425 is defined as the T-associated allele of DYF371. For example, DNAFP might report the results for DYF371 as “10c-12t-13c-13c”, from which the DYS425 value is shown to be 12.

The repeat value for DYS452 is reported differently by various companies. The marker consists of one continuous repeating TATAC structure of about 12 repeats, plus 19 additional contiguous units made up of CATAC, TGTAC, or TATAC units. These 19 repeats are normally invariant. Some companies (DNAH and RG) report only the main (variable) repeat value of TATAC (12 in the above example), while others (SMGF, DNAFP) add the other 19 repeats as well for a total of 31 (and this is also the ISFG/NIST-standard nomenclature). We will use the latter notation here.

There are apparently no differences in nomenclature onw are c re a0.n the turethe 5RGa exa

ainpeateach0.8(allt(Thadja)TJ13.29341 0 TD09.001 T

haplogroup designations or ambiguous results from the Y-Haplogroup Predictor, other steps were taken to determine the haplogroup, such as the origin of the family in the OA record.

Until recently, the markers DYS452 and DYS446 were tested only by Sorenson Genetics and its resellers DNAH and RG. Now these markers are also available from DNAFP. Since the SMGF database covers both of these markers, it was used as the primary source of information on these markers.

Candidate Haplogroup G haplotypes were extracted from the SMGF database using somewhat different search criteria[4] from those used for the OA database. Candidate haplotypes were tested using the Haplogroup Predictor Program (Athey 2005) and only those with a score exceeding 50 for Haplogroup G were used. Multiple haplotypes with the same surname listed were deleted, retaining only one haplotype per surname (except where the haplotypes were clearly unrelated).

The marker DYF399S1 is only available from DNAFP, and none of the public databases (except DNAFP's own Y-Match) currently accept data on this markY-Matchat different

,

For members of Haplogroup G2, they are reporting a value that is four repeats less than what is actually present. This is a good reason for using the ISFG/NIST standard nomenclature.

Allele frequencies on DYS452 for the most common European haplogroups are shown in **Table 2**. The values for Haplogroup G2 are smaller than for most haplogroups.

The limited data for GxG2 suggests that the deletion event in DYS452 occurred in a Haplogroup G2 individual or else was present in the founder of G2. The value of 27 on this marker for the G2 individual from a tribal area of India supports the idea that the deletion occurred in a person who was already G2, or that it occurred very early in the history of G2. Probably, the

,

Table 4 Allele Frequencies for DYF399S1 in Haplogroup G

DYS399S1a			DYS399S1b			DYS399S1c			DYS399S1d		
Repeats	Count	Freq.	Repeats	Count	Freq.	Repeats	Count	Freq.	Repeats	Count	Freq.
17.2	4	.571	(missing)	1	.143	21	2	.286	(missing)	5	.714
18.2	2	.286	20.1	5	.714	22	3	.429	24	1	.143
19.2	1	.143	21.1	0		23	0		25	1	.143
			22.1	1	.143	24	2	.286			

where the lower case letters are part of a countable repeat motif and the upper case letters are "extra" bases (10 of them in the above example). This example would be scored as 18.10 (or 18.0 in the short notation).

There are as yet only a few results available for DYF399S1, but those available for Haplogroup G (all but one are G2 and the remaining one is GxG2) are shown in **Table 4**. Beside the odd structure for the shortest allele in Haplogroup G2, the whole number of repeats in the shortest allele is lower than in Haplogroups R1b, I, and J, although one example in Haplogroup I is the lowest so far found (16). Each G person represented in Table 4 has one allele with the "half" repeat (a .2 following the main number), one allele with a .1 following the main number, and one allele with a whole number. Even though the three alleles are reported in numerical order, the alleles can be distinguished for members of Haplogroup G. One person in the table had four allele values (a member of G2), apparently representing a doubling of the allele with the whole number of repeats.

Discussion

The modal value for DYS452 for Haplogroup G2 was found to be 26, lower than for any other haplogroup. The modal value for DYS446 in Haplogroup G2, in contrast, is the highest for any haplogroup. Therefore, the difference in values on these two markers would be particularly diagnostic of Haplogroup G2. The difference (DYS452 – DYS446) will typically have a value of 11 or less in Haplogroup G2, but a value of 17 or higher in other haplogroups. Interestingly, members of Haplogroup GxG2 appear to have similarly large values on DYS446, but do not have low values on DYS452. Therefore, DYS452 can serve to distinguish G2 from other parts of G.

The shortest allele of the marker, DYG399S1 has a small number of whole repeats in Haplogroup G, and also has a fractional repeat value that has only been found so far in Haplogroup G.

Members of Haplogroup G are fortunate to have several Y-STR markers that are either diagnostic or strongly suggestive of membership in G. However, these diagnostic markers are not tested as often as other STRs. With these markers now generally available, they will be of value in predicting Haplogroup G.

Electronic-Database Information

https://home.comcast.net/~whitathey/predictorinstr.htm
Haplogroup Predictor Program

http://www.oxfordancestors.com/index.html
Oxford Ancestors Database

http://www.ysearch.org
Y-Search Y-STR Public Database

http://www.ybase.org
Y-Base Y-STR Public Database

http://www.smgf.org
Database of Sorenson Molecular Genetics Foundation

http://www.dna-fingerprint.com/modules.php?op=modload&name=ymatch
Y-Match Y-STR Public Database

References

Athey TW (2005) Haplogroup prediction using an allele-frequency approach. J Genetic Genealogy, 1:1-7.

Behar DM, Garrigan D, Kaplan ME, Mobasher Z, Rosengarten D, Karafet TM, Quintana-Murci L, Ostrer H, Skorecki K, Hammer MF (2004) Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum Genet, 114:354-365.

Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, Foren Sci Int, 129:10-24.

Henson G (2005) DYF399S1: A unique three-copy short tandem repeat on the human Y chromosome. J Genetic Genealogy, 1:8-11.

Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700-705.