Journal of Genetic Genealogy 1:35-39, 2005

Pitfalls in Determinations of Y Haplogroup F*

T. Whit Athey

Evidence from several case studies is presented showing that when an individual is determined to belong to the Y Haplogroup F*, based upon negative results for the defining SNPs of the major sub-haplogroups of F, this result is usually not correct when the individual's Y-STR values clearly suggest membership in one of those sub-haplogroups. These case studies include examples that were ultimately proven to belong to Haplogroup G, Haplogroup I, and Haplogroup J. The negative results on SNP markers M201, P19, and 12f2.1, which in part led to the conclusion that these individuals were F*, were shown to be due simply to lab errors. The evidence for one of the cases appears to support the F* determination.

Introduction

Y-chromosome Haplogroup F is a large "macro-haplogroup" that includes much of the world's population. Nearly all of that population is in further derived sub-haplogroups defined by downstream single-nucleotide polymorphisms (SNPs). Only in India have significant numbers of people been reported to be in the root of Haplogroup F, that is, in Haplogroup F, but not having any downstream SNP mutations (Kivisild et al. 2003) defining the present Y phylogenetic tree. The haplogroup for such individuals would properly be called Haplogroup F*.

Small numbers of men in Haplogroup F* were reported in the study of Iberian Y chromosomes by Flores et al (2004). In two of the north Portuguese populations that were studied, a total of three F* individuals were found, representing 0.5% of the overall study population. Possibly, the presence of these individuals resulted from admixture—Portugal had significant contacts with India about 500 years ago.

The SNP status for Haplogroup F* individuals would be M89+ (or the apparently redundant P14+, defining Haplogroup F), but M201- (not in Haplogroup G), M69- (not in Haplogroup H), M170- (or redundantly, P19- or M258-, not in Haplogroup I), 12f2.1- (or redundantly M304-, not in Haplogroup J), and M9- (not in macro-Haplogroup K). One additional subgroup of Haplogroup F may be Haplogroup F1, but the status of this group, possibly defined by SNPs P91 and/or P104, is presently uncertain. If Haplogroup F1 exists, it is likely to be many orders of magnitude smaller than the others mentioned above, and it is probably safe to ignore it.

Received: November 10, 2005, accepted December 17, 2005

Address for correspondence: wathey@hprg.com

Recently, several different individuals were tested for a series of SNPs to determine their haplogroup by a commercial testing company. Four of these individuals had Y-STR data that suggested that they belonged to Haplogroups G, I, or J. The SNP series for all four showed that they were M89+ or P14+, so that all three belonged somewhere within macro-Haplogroup F. All four individuals, however, were found to be negative for one of the defining SNPs for each of the major haplogroups within Haplogroup F. That is, they were found negative for M69 (H), M201 (G), P19 (I), 12f2.1 (J), and M9 (K), and they were declared by the testing company to be in Haplogroup F*. A fifth individual, ordering individual SNP tests on his own, had the same SNP results, even though his Y-STR values suggested membership in Haplogroup G.

Since there would be no reason for the Y-STR values of an F* individual to resemble those of any of the major subgroups, further SNP testing of these five individuals, along with a cousin of two of them, was carried out in the present study. It was first assumed that the original SNP results were correct, and that possibly back mutations had occurred, so the initial tests were on SNP markers downstream from the defining major haplogroup markers. In every case except one, a downsteam marker was found to be positive and a subsequent retest of the corresponding major haplogroup SNP (M201, 12f2.1, or M170) was positive. In every case except one, Y-STR values consistently pointed to the same haplogroup that was confirmed by the SNP tests.

These previous SNP results seemed quite inconsistent with the Y-STR data and seemed likely to be explained by one of three possibilities: (1) a back-mutation had occurred on the defining SNP for a major haplogroup in each case, (2) the original negative SNP test results were incorrect, or (3) some amazing coincidences had occurred in the mutations on the short tandem repeats to make these individuals appear to be in a haplogroup where

they did not belong. To decide between these three possibilities, we tested several SNPs downstream from the major SNP markers that had previously shown negative results. When these were all positive, we also rechecked those SNPs previously found negative.

In at least one case, the original negative SNP result had been found at two independent labs, so there was an expectation that at least some of these cases would be found to result from back mutations. However, in every case where the Y-STR values strongly suggested a particular haplogroup, the anomalous results were ultimately found to be the result of lab errors.

The company that assigned the "Haplogroup F*" designation to four of the study participants should have known that something was wrong from the fact that the Y-STR values were strongly suggesting membership in a particular haplogroup. Aside from that, it should have been suspicious because of the apparent rarity of persons who are actually in Haplogroup F*, particularly persons who have backgrounds in northwestern Europe.

The results of the present study demonstrate that commercial testing companies should be very wary of concluding that a customer is in Haplogroup F*. When faced with the possibility of an F* case, the STR values for the customer should be used as a guide to rerun one or more of the original SNP tests, and if still negative, then a SNP test should be run for the most likely downstream subgroup. Only after these confirming results are available should an F* assignment be made.

Methods

Subjects were chosen from five different surname projects (Athey, Bell, Boyette, Owen, and Power surnames) where SNP tests on the subjects themselves, or on a related member of the same project, had indicated that the individuals were in Haplogroup F*. In addition to these five subjects, one additional related participant from the Athey and Owen surname projects were added to the study as a check on the SNP status of the two primary subjects. The subject from the Boyette surname project was included even though his Y-STR values did not strongly suggest a particular haplogroup. All subjects had tests for either 25 or 37 Y-STR values, all carried out at Family Tree DNA (FTDNA, Houston, TX). The two subjects from the Athey surname project are cousins who share a common ancestor who lived from 1642 to 1709, and their 37 Y-STR values match 36/37 on an unusual haplotype. The two subjects from the Owen surname project do not know their exact relationship, but they are undoubtedly cousins and their Y-STR values match exactly on an unusual 37-marker haplotype. They share a common ancestor who probably lived from 1676 to 1767.

All of the Y-STR values were obtained through the respective surname projects, with the permission of the participants, and were known prior to the present study with the exception of markers 26-37 for subject J-105 (which were obtained during the study).

SNP tests on some of the seven subjects had previously been carried out at FTDNA and one subject had been tested by Trace Genetics (Richmond, CA) using methods that have not been described in detail by the companies. Both state that they follow published methods.

For the present study, SNP testing was carried out by Ethnoancestry (Cyprus, CA) for all markers except 12f2.1, which was tested at FTDNA. At Ethnoancestry, Y chromosome SNPs were amplified by PCR with standard primers giving products from 200 to 500 bp in length. PCR products were then sequenced using dye terminator chemistry with electrophoresis on a capillary ABI sequencer. Alleles were called in Sequencher by alignment with chromosomes of known allelic state (positive and negative controls).

Results

The Y-STR values for all subjects were available from participation in surname projects and are shown in **Table 1**. Estimated haplogroups were scored for each Y-STR haplotype using an allele frequency approach (Athey, 2005). **Table 2** shows the haplogroup scores for 11 haplogroups that occur in Europe (some occur only rarely). In six of the seven cases, one haplogroup received significantly higher scores than the others, strongly suggesting that those six subjects were members of that haplogroup. For the seventh subject the highest scores were for Haplogroups J2 and G, but the scores were fairly low and were not definitive.

Prior to the present study, an initial series of SNP tests on subjects G102, X103,[1] J104, I106, and I107,[1] resulted in the conclusion by the testing company (FTDNA, except G102) that these individuals were in Haplogroup F*. However, because of the rarity of F*, especially in northwest Europe, and the strong indications of the STR values, further SNP tests for haplogroups suggested by the Y-STR values were conducted in the present study. A summary of the SNP results for all of the subjects obtained prior to the present study is presented in **Table 3**.

The haplogroup scores were highest on Haplogroup G for subjects G101 and G102, but G102's M201 status

[1] For subjects X103 and I107, it was actually another individual with matching Y-STR values in the same surname project who had been SNP tested. These two individuals, along with other matching members of the same surname project, likely shared a common SNP status.

had been found negative (ancestral) at two independent labs, so these were first tested for the SNP that defines the most common subclade (G2) of Haplogroup G, namely P15. Both G101 and G102 were found to be P15+ and so are in Haplogroup G2. All (next-lower-level) downstream SNPs within Haplogroup G2 were found to be negative for these two subjects, specifically P16 and M286. A retest of M201 for Subject G102, and an initial test of M201 for Subject G101, showed both to be M201+. Therefore, the initial M201- results for G102 was shown to be incorrect. See **Figure 1**.

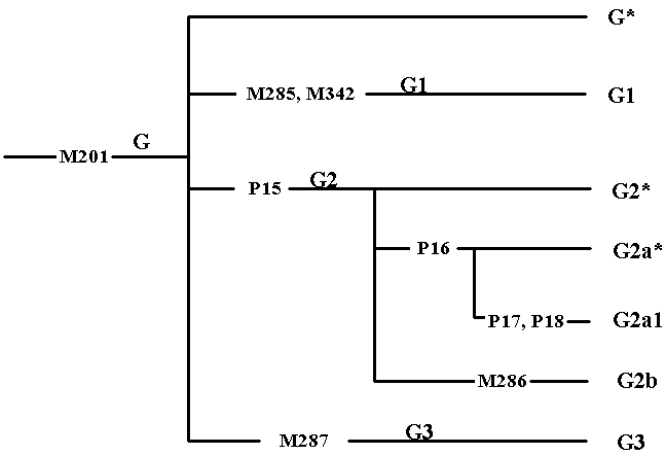

Figure 1 Phylogenetic Chart for Haplogroup G (simplified)

Subject J105 was tested previously for the SNP, M172 by FTDNA, and was found to be positive, so the same SNP was first tested in Subject J104. Subject J105's situation was opposite from all of the other subjects in that this subject had been tested only for M172, a marker downstream of the SNP defining a major haplogroup (J), and had been found to be positive, but he had not been tested for a defining SNP for Haplogroup J. Subjects J104 and J105 were also tested for the upstream marker M304, and both were found to be M304+ (in Haplogroup J). Subject J104 was also found to be M172+ (in Haplogroup J2). See **Figure 2**.

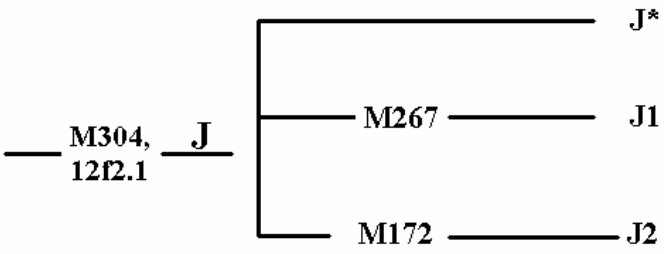

Figure 2 Phylogenetic Chart for Haplogroup J (simplified)

Since Subject J104 had previously been found to be 12f2.1- at FTDNA, and since Ethnoancestry does not test for 12f2.1, tests for this SNP for Subjects J104 and J105 were ordered at FTDNA. Both were found to be 12f2.1+, confirming that the original result of 12f2.1- for subject J105 had been incorrect.

Subjects I106 and I107 were tested for the SNPs that define the subclades of Haplogroup I, based on the haplogroup scores for their Y-STR values. See Figure 3. Subject I106, whose haplogroup scores suggested membership in Haplogroup I1c, was found to be M223+, confirming the prediction based upon STR values. He was also found to be M170+ and P19+, showing that the earlier P19- result had been incorrect. The haplogroup prediction for subject I107 was clearly I1a, and he was not surprisingly found to be M253+. He was also found to be M170+ and P19+, demonstrating that his P19- result had also been in error. Both of these subjects were also found to be P38+. See **Figure 3**.

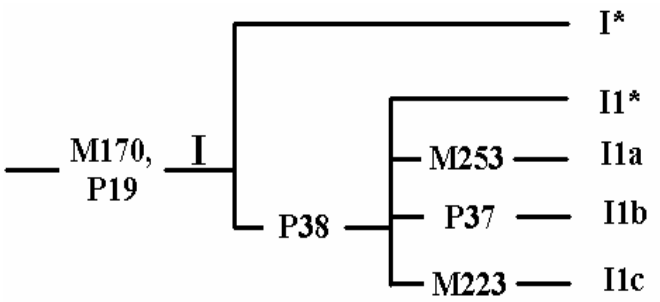

Figure 3 Phylogenetic Chart for Haplogroup I (simplified)

Subject X103 had haplogroup scores that weakly indicated that he might be in either Haplogroup J or G. A cousin of this subject had already been tested for M201 (G) and P15 (G2) and was found negative on both markers, so he was tested first for M285 (G1). However, he was M285- and also M287- (not in G3). A repeat of the P15 test showed a negative result, as did tests for two downstream markers within G2: P16- (G2a) and P286- (G2b). Next M201 and M304 were tested for this subject and he was found to be M201- and M304-. Therefore, it appears that this subject could well be in Haplogroup F*.

The results for all of the SNP tests from the present study are shown in **Table 4**.

Table 1 – Y-STR Values* for All Subjects

ID	393	390	19	391	385a	385b	426	388	439	389-1	392	389-2	458	459a	459b	455	454	447	437	448	449	464a	464b	464c	464d	460	GATA H4	YCAIIa	YCAIIb	456	607	576	570	CDY a	CDY b	442	438
G101	14	22	16	10	11	13	11	12	12	14	11	17	16	9	9	11	11	22	16	21	31	12	12	12	13	10	10	20	20	15	15	20	16	39	39	11	10
G102	14	22	16	10	11	13	11	12	12	14	11	17	16	9	9	11	11	22	16	21	31	12	12	12	13	10	10	20	20	15	15	20	17	39	39	11	10
X103	14	22	14	10	12	13	11	12	13	13	11	17	17	8	9	10	11	23	15	19	30	11	12	15	16	10	10	17	17	15	13	18	18	36	37	12	10
J104	12	23	14	10	12	13	11	15	12	14	12	17	15	9	9	11	11	26	15	20	32	12	13	15	16	11	10	22	22	17	14	20	19	34	34	11	9
J105	12	23	14	10	12	13	11	15	12	14	12	17	15	9	9	11	11	26	15	20	32	12	13	15	16	12	10	22	22	17	14	20	20	34	34	11	9
I106	15	24	15	10	15	15	11	13	13	14	12	17	15	8	10	11	11	25	14	21	27	11	14	14	15												
I107	15	23	14	10	13	14	11	14	12	12	11	16	16	8	9	8	11	23	16	20	29	12	14	15	17	10	10	19	21	14	14	15	18	33	36	13	10

* All subjects were tested by FTDNA and the results are presented in the same order as in an FTDNA report.

Table 2 Haplogroup Predictor Scores

ID	Goodness of Fit Scores for Eleven Haplogroups										
	E3a	**E3b**	**G**	**I1a**	**I1b**	**I1c/ I2**	**J2**	**N**	**Q**	**R1a**	**R1b**
G101	5	5	**43**	6	17	9	16	2	10	10	1
G102	5	5	**43**	6	17	9	16	2	10	10	1
X103	7	9	**21**	13	12	9	**30**	4	9	4	4
J104	6	17	24	10	27	18	**67**	6	26	11	7
J105	5	16	23	11	28	17	**67**	7	26	11	7
I106	19	15	25	7	46	**72**	31	4	20	7	5
I107	15	15	23	**64**	22	18	43	3	16	7	5

Table 3 – Previous SNP Results for All Subjects

ID	SNP (Haplogroup That It Tests), Lab*, Test Result (+ or -)									
	M89 (F)	**P14 (F)**	**M201 (G)**	**P15 (G2)**	**M69 (H)**	**M170 (I)**	**P19 (I)**	**12f2.1 (J)**	**M172 (J2)**	**M9 (K)**
G101										
G102	EA+	FT+	FT- TG-		EA-	TG-	FT-		EA-	EA- FT-
Cousin of X103	FT+		FT-	FT-	FT-		FT-	FT-		FT-
J104		FT+	FT-		FT-		FT-	FT-		FT-
J105									FT+	
I106			FT-		FT-		FT-	FT- (M304-)		FT-
Cousin of I107	FT+		FT-		FT-		FT-	FT-		FT-

* FT=Family Tree DNA, TG=Trace Genetics, EA=Ethnoancestry

Table 4 – SNP Results for All Subjects, This Study*

ID	SNP (+ or -)				
G101	M201+	P15+			
G102	M201+	P15+	P16-	M286-	
X103	M201-	P15-	M285- M287-	M286-	M304-
J104	M304+	M172+	M267-	*12f2.1+*	M47- M158- M67-
J105	M304+			*12f2.1+*	
I106	M170+ P19+	P38+	M223+	M253- P37-	M284-
I107	M170+ P19+	P38+	P253+ P30+	M223- P37-	P40+ M227-

* All results are from Ethnoancestry except those in italic font, which were run at FTDNA.

Conclusion

Genetic genealogy companies should be very wary of labeling a customer as a member of Haplogroup F* when that customer's Y-STR values strongly suggest membership in a particular major subgroup of F. The same could be said concerning labeling a customer with any haplogroup when his STR values suggest a different haplogroup. A truly F* individual would have no reason to have STR values that are similar to those of Haplogroups G, H, I, or J. A more likely explanation is a false negative result on one of the SNP tests. In such cases the company should first retest the SNP marker for the haplogroup suggested by the STR values, and if still negative, test one or more downstream markers in the probable haplogroup.

When a subject has the SNP status suggesting F*, and his Y-STR values do not resemble any of the subgroups of F, then F* is a real possibility, though even here a retest of each SNP would be advisable.

Added Note

At about the same time that the current study was completed, FTDNA, on its own initiative, retested M201 for subject G102 and this time (16 months after the M201- result) he was found M201+. FTDNA also found subject G102 and another participant in the Athey surname project to be P15+.

As a result of the tests of 12f2.1 that were ordered from FTDNA as a part of this study, FTDNA has removed the F* designation for Participant J104 and now shows his haplogroup as J2. When informed of the results on subject I106, FTDNA stated that this subject would be retested as well.

Electronic-Database Information

http://www.hprg.com/hapest5/
haplogroup predictor

References

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Molecular Phylogenetics and Evolution, 28:458-72.

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

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148.

Flores C, Maca-Meyer N, Gonzales AM, Oefner PJ, Shen P, Perez JA, Rojas A, Laruga JM, Underhill PA (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. Eur J Hum Genet, 12:855-863.

Jobling MA, Tyler-Smith C (2003). The human Y chromosome: an evolutionary marker comes of age. Nature Reviews—Genetics, 4:598-812.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72: 313-332.