A similar-sounding surnames sequel: haplogroup R-FT70038

by John Maury Branan, Jr. ✉

Genealogist, Billerica, MA, USA

Journal of the Fitzpatrick Clan Society 2023, 4, 11-44

Appendices

1. Y-STR mutations matrix
2. Autosomal match matrix
3. Modifying the Y-SNP/STR haplotree by optimum use of autosomal match data
4. Monte Carlo simulation – Y-STR mutations
5. Monte Carlo simulation – number of autosomal matches

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Appendix 1: Y-STR mutation summary matrix

	Project Participant - BRAN																				Mutations
STR Marker	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**Total Derived**
DYS447	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	1
DYS449	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1
DYS464a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	1
Y-GATA-H4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DYS607	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1
DYS576	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3
DYS570	0	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	2
CDYa	0	0	0	-1	0	-1	0	0	0	0	0	0	0	-1	0	0	-2	0	0	0	0	4
CDYb	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DYS438	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
DYS413a	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1
DYS534	1	0	0	0	0	0	0	1	1	1	0	0	0	0	-1	0	0	0	1	0	1	7
DYS446	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
DYS710	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	3
DYS540	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	2
DYS714	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DYS533	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	3
Y-GATA-A10	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DYS712	-1	0	0	0	0	-1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3
DYS650	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1
DYS532	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DYS635	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
DYS587	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
DYS497	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2	0	1
Total Derived	5	1	2	2	2	5	2	2	1	2	0	1	1	4	2	3	4	1	1	2	3	46

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Appendix 2: autosomal match matrix

Caron Brannon — Thomas Brannan **FTC4333**; James Brannon **FT101136**

Kenyon Branan — Caswell Harris Branan **FTB27810**

Unknown Brannon — **Y10443**

	BRAN1	BRAN2	BRAN3	BRAN4	BRAN5	BRAN20	BRAN6	BRAN7	BRAN21	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14	BRAN15	BRAN17	BRAN19
BRAN1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BRAN2	0		0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BRAN3	0	0		0	0	0	28	0	0	0	0	13	0	0	0	0	0	0	0
BRAN4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BRAN5	0	22	0	0		0	0	0	0	0	10	11	0	0	0	0	0	0	0
BRAN20	0	0	0	0	0		20	0	0	0	0	0	0	0	0	26	16	0	0
BRAN6	0	0	28	0	0	20		18	19	0	0	0	0	0	0	0	0	0	0
BRAN7	0	0	0	0	0	0	18		0	19	0	0	0	0	0	0	0	0	0
BRAN21	0	0	0	0	0	0	19	0		167	120	153	65	186	0	50	0	0	0
BRAN8	0	0	0	0	0	0	0	19	167		393	347	0	28	0	0	0	0	0
BRAN9	0	0	0	0	10	0	0	0	120	393		1945	52	93	0	13	0	0	0
BRAN10	0	0	13	0	11	0	0	0	153	347	1945		0	17	0	0	0	0	0
BRAN11	0	0	0	0	0	0	0	0	65	0	52	0		50	0	57	0	0	0
BRAN12	0	0	0	0	0	0	0	0	186	28	93	17	50		14	121	0	0	0
BRAN13	0	0	0	0	0	0	0	0	0	0	0	0	0	14		11	0	0	0
BRAN14	0	0	0	0	0	26	0	0	50	0	13	0	57	121	11		0	0	0
BRAN15	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0		0	0
BRAN17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
BRAN19	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Appendix 3: Modifying the Y-SNP/STR haplotree by optimum use of autosomal match data

In Figure 8C we presented a Haplotree which modified the Y-DNA Haplotree via the use of autosomal match data. In this Appendix we describe the approach and then discuss merits and shortcomings of that optimization of the autosomal match data. We will discuss the algorithm as applied to BRAN8 through BRAN14, which is the most complex network in the set. The other sets were optimized through inspection as previously described. For BRAN8 through BRAN14:

1. We start with the autosomal match matrix, segment length of shared DNA in cM from Figure 3.
2. We translate the shared segment lengths to most likely relationships from the data in Figure 2.
3. We use Donnelly's formulas for calculating the relevant matched pair relationships in our set and rewrite the matrix using that transformation[28]. This transformation has the utility of linearizing segment length versus number of separating generations.
   a. $k = 2s + t$ for sth cousins t times removed
   b. $k = n + 1$ for (great)^n uncle (uncle n times removed)
4. The resulting matrix is defined as the optimum autosomal set A, and is described the sum of k values above the diagonal:

$$S^{\Delta} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f(i,j)$$

5. We seek a haplotree with a thusly transformed matrix, B, with k approaching the optimum autosomal set's k value. We define tree error for matrix B:

$$\text{Tree Error} = k_B - k_A$$

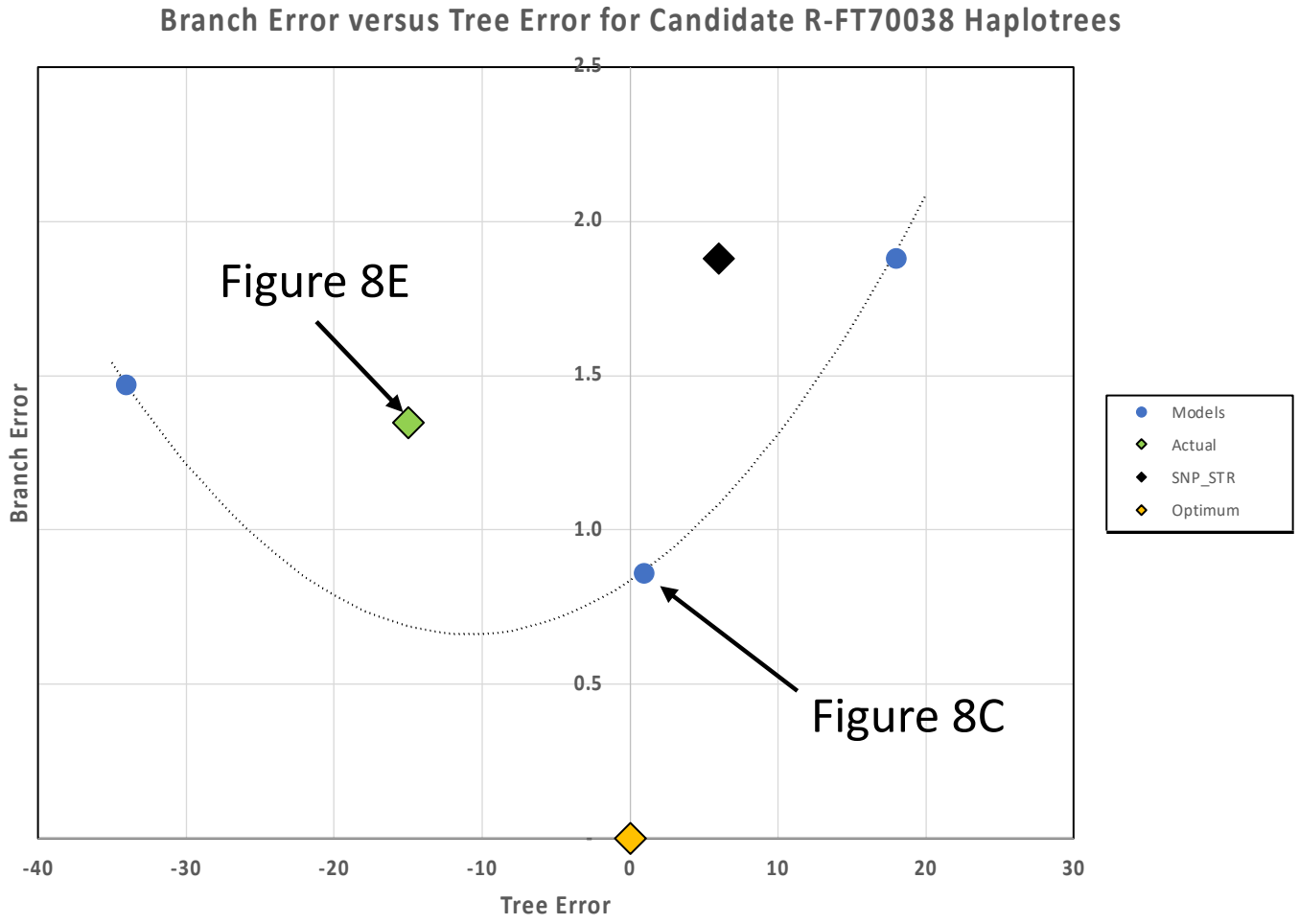
6. We constrain the solution sets of haplotrees to adhere to the Y-SNP/STR tree, i.e., the para-haplotree members BRAN 12-14 will always have a more distant MRCA than the R-FT27810 members.
7. We also seek a haplotree with each autosomal pair k value as close as possible to the corresponding k value in the optimum autosomal set. We define Branch Error for the matrix B relative to matrix A, both n x n sized matrices:

$$\text{Branch Error} = \frac{\sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \left(|k_{A(i,j)} - k_{B(i,j)}|\right)}{n^2}$$

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Figure A is a graph of the resulting branch error versus tree error for five candidate trees. The Y-SNP/STR haplotree from Figure 8B is farther away from the optimum haplotree A than the tree constructed to optimize autosomal match results. The overall k-value of the Y-SNP/STR haplotree matches the optimum tree's total k-value quite well. However, many of its matched pairs are high above or high below the optimum case, driving large deltas in individual pairs. One of the three constructed trees is somewhat closer to the optimum tree than the actual tree based on Figure 8E, illustrating the natural variability. The full algorithm, including development of the tree and branch error values for the Y-SNP/STR haplotree is shown in Figure B and Figure C as an example.

Figure A: branch error versus tree error – autosomal data from the James Branan descendants



Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

The merit of this approach is that haplotree candidates can be compared objectively versus a standard on two dimensions: tree error, which is overall coverage of the space, roughly correlating with number of generations, and branch error, which measures how well the tree fits each and every matched pair. There are at least two shortcomings of this method, both associated with the definition of the optimum tree.

The first shortcoming is that the declaration of the optimum tree is based solely on the shared centimorgan project's statistics which are silent on the impact of 'no match' findings. The second shortcoming is that transforming a shared centimorgan finding into a specific relationship has ambiguity (e.g., should one assign a 13 cM match to a 5C3R or a 6C2R relationship?) and also ignores the variation reported in the Bettinger data.

British statistician George Box stated, 'all models are wrong, but some are useful.' The usefulness of this model is to suggest that eleven generations in the haplotree, not eight generations best fits the available data.

Figure B: Development of tree error for Y-SNP/STR tree

Development of Optimum Autosomal Tree k-value

Shared Autosomal segments, centimorgans

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8		393	347	0	28	0	0
BRAN9	393		1945	52	93	0	13
BRAN10	347	1945		0	17	0	0
BRAN11	0	52	0		50	0	57
BRAN12	28	93	17	50		14	121
BRAN13	0	0	0	0	14		11
BRAN14	0	13	0	57	121	11	

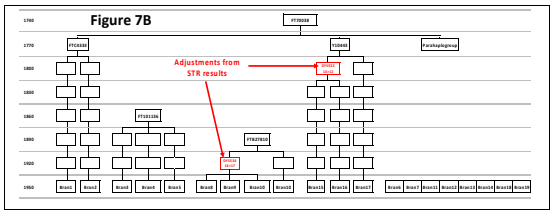
Most likely relationship

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8		1C1R	!C!R		4C1R		
BRAN9	1C1R		U	3C1R	3C		6C2R
BRAN10	!C!R	U			6C		
BRAN11		3C1R			3C1R		3C1R
BRAN12	4C1R	3C	6C	3C1R		7C	2C1R
BRAN13				0	7C		8C
BRAN14		6C2R		3C1R	2C1R	8C	

Relatedness, k-value

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14	K Total
BRAN8		3	3		9			15
BRAN9	3		1	7	6		14	28
BRAN10	3	1			12			12
BRAN11		7			7		7	14
BRAN12	9	6	12	7		14	5	19
BRAN13					14		16	16
BRAN14		14		7	5	16		
								104

Development Branch Error for SNP/STR based Haplotree



SNP/STR Matrix

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8		B	B		6C		
BRAN9	B		B	1C	6C		6C
BRAN10	B	B			6C		
BRAN11		1C			6C		6C
BRAN12	6C	6C	6C	6C		6C	6C
BRAN13					6C		6C
BRAN14		6C		6C	6C	6C	

K - Figure 7B

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14	K Total
BRAN8		0	0		12			12
BRAN9	0		0	2	12		12	26
BRAN10	0	0			12			12
BRAN11		2			12		12	24
BRAN12	12	12	12	12		12	12	24
BRAN13					12		12	12
BRAN14		12		12	12	12		
								110

Tree Error = 110 – 104 = 6

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Figure C – Development of branch error for Y-SNP/STR Tree

$$\frac{\sum_{i=1}^{i=n}\sum_{j=1}^{j=n}\left(\left|k_{A(i,j)}-k_{B(i,j)}\right|\right)}{n^2}$$

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8		3	3		9		
BRAN9	3		1	7	6		14
BRAN10	3	1			12		
BRAN11		7			7		7
BRAN12	9	6	12	7		14	5
BRAN13					14		16
BRAN14		14		7	5	16	

−

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8		0	0		12		
BRAN9	0		0	2	12		12
BRAN10	0	0			12		
BRAN11		2			12		12
BRAN12	12	12	12	12		12	12
BRAN13					12		12
BRAN14		12		12	12	12	

$(7)^2$ =

	BRAN8	BRAN9	BRAN10	BRAN11	BRAN12	BRAN13	BRAN14
BRAN8	0	3	3	0	3	0	0
BRAN9	3	0	1	5	6	0	2
BRAN10	3	1	0	0	0	0	0
BRAN11	0	5	0	0	5	0	5
BRAN12	3	6	0	5	0	2	7
BRAN13	0	0	0	0	2	0	4
BRAN14	0	2	0	5	7	4	0

49

$$= \frac{(3+3+3+3+1+5+6+2+3+1+5+5+5+3+6+5+2+7+2+4+2+5+7+4)}{49}$$

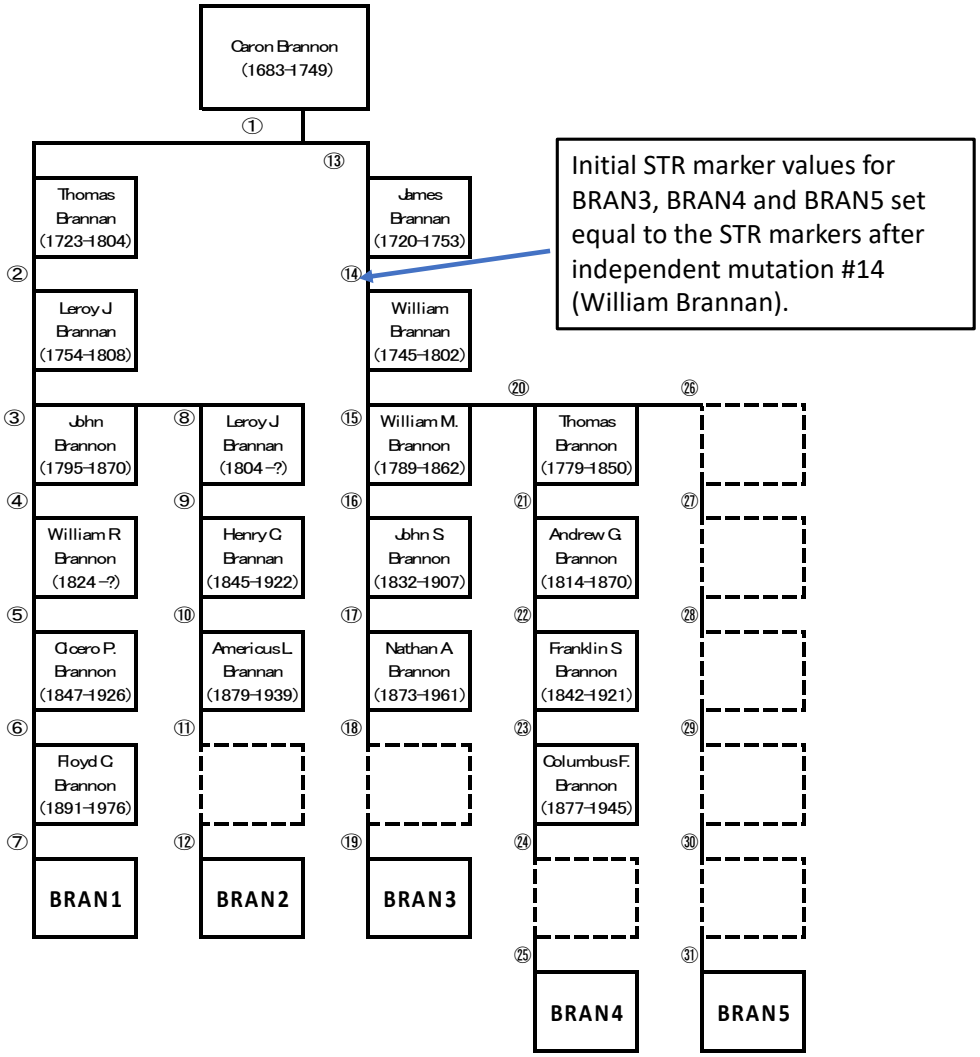
$$= 1.88$$

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Appendix 4: Y-STR Monte Carlo simulation

Each generation represents know mutation opportunities on each of 111 STR markers. Figure D illustrates a haplotree with 31 mutation opportunities. The most likely haplotree in Figure 13 has 125 total mutation opportunities, so each run of the simulation calculates 125 generationally sequential mutation opportunities for each of the 111 markers. A 2000-run simulation was performed in Microsoft Excel using the random number generator function, RAND(), for each mutation opportunity. In Figure 20, as an example, the BRAN 1 and BRAN3 lines are simulated with initial values for each Y-STR marker in the initial generation. All other BRAN simulations, BRAN2, BRAN4 and BRAN5 take their initial values from the appropriate generation of the simulations for BRAN1 and BRAN2. In our models we used 'zero' as the modal value for the Y-STR marker.

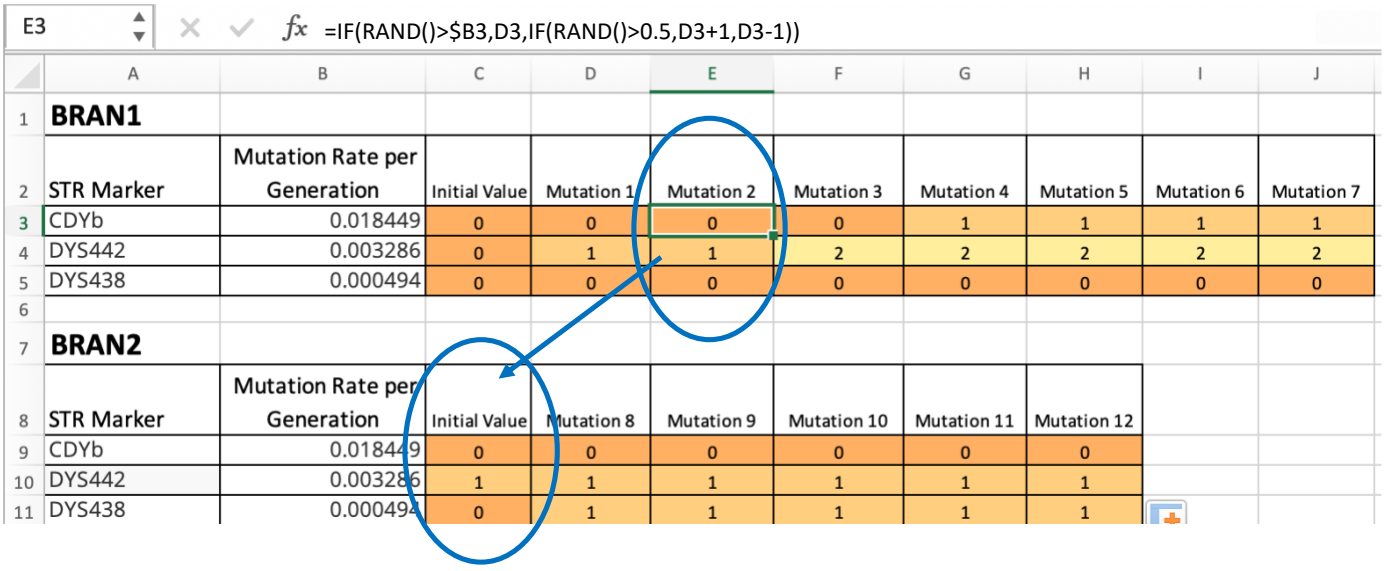
Figure D – Technique for Simulating IBD Mutations



Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Figure E illustrates in more detail the simple simulation based on a small portion of the tree shown in Figure D. The mutation formula for cell E3 references the mutation rate for the Y-STR Marker CDYb (0.018449 per generation) and returns cell D3 marker value for all random numbers > 0.018449. If the random number generated is <0.018449 then the equation increases or decreases the Y-STR marker value by 1. In all cases, BRAN2's initial values for Y-STR markers are set equal to BRAN1's mutation #2 result (Leroy J Brannan). In this example for CDYb, BRAN1 has a mutation in mutation #4, but BRAN2 does not. For DYS442, BRAN1 has a mutation in mutation #1 and mutation #3 – the value of the marker after mutation #2 is 1 which is passed along to BRAN2 as the initial value of the marker. Note that BRAN1's mutation #3 change has no impact on BRAN2. For DYS438, BRAN2's initial value changes in mutation #8, while BRAN1 has no mutation.

Figure E – Equations for Monte Carlo simulation of Y-STR mutations in R-FT70038 haplotree



	A	B	C	D	E	F	G	H	I	J
1	**BRAN1**									
2	STR Marker	Mutation Rate per Generation	Initial Value	Mutation 1	Mutation 2	Mutation 3	Mutation 4	Mutation 5	Mutation 6	Mutation 7
3	CDYb	0.018449	0	0	0	0	1	1	1	1
4	DYS442	0.003286	0	1	1	2	2	2	2	2
5	DYS438	0.000494	0	0	0	0	0	0	0	0
6										
7	**BRAN2**									
8	STR Marker	Mutation Rate per Generation	Initial Value	Mutation 8	Mutation 9	Mutation 10	Mutation 11	Mutation 12		
9	CDYb	0.018449	0	0	0	0	0	0		
10	DYS442	0.003286	1	1	1	1	1	1		
11	DYS438	0.000494	0	1	1	1	1	1		

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023

Appendix 5: Autosomal DNA Monte Carlo simulation

For any pair of individuals, there is a reported probability of no detectable shared DNA, as shown in Figure 2. Figure F shows as an example a match simulation created for BRAN8 versus a subset of members of the R-FT70038 Haplotree. As in our Y-STR Monte Carlo simulations reported in Appendix 4, we used the random number generator function in Microsoft Excel to determine if shared DNA would be detected for each pair in each run. For the ten runs shown for BRAN8 in Figure F, the number of autosomal matches ranged from three to five. The formula for the indicated cell E126 is visible as reference. This approach was expanded to encompass all 19 autosomal test results over 2000 runs of the model to generate the statistics reported in Figure 11 in the main text.

Figure F: Autosomal DNA assessment of number of matches for haplogroup members

E126 =IF(RAND()<$D126,0,1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
32	**Match 1**	**Match 2**	**Relationship**	**Probability of No Detectable DNA Relationship**	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 7**	**Run 8**	**Run 9**	**Run 10**
126	BRAN8	BRAN9	1st cousin once removed	0.0%	1	1	1	1	1	1	1	1	1	1
127	BRAN8	BRAN10	2nd cousin	0.0%	1	1	1	1	1	1	1	1	1	1
128	BRAN8	BRAN11	3rd cousin	2.3%	1	1	1	1	1	1	1	1	1	1
129	BRAN8	BRAN12	4th cousin	30.7%	0	1	1	1	1	1	0	1	0	1
130	BRAN8	BRAN13	5th cousin	69.8%	0	0	1	1	0	1	1	0	1	0
131	BRAN8	BRAN14	4th cousin once removed	52.0%	0	1	0	0	0	0	1	0	0	1
132	BRAN8	BRAN15	7th cousin once removed	98.4%	0	0	0	0	0	0	0	0	0	0
133	BRAN8	BRAN17	8th cousin	99.2%	0	0	0	0	0	0	0	0	0	0
134	BRAN8	BRAN19	7th cousin once removed	98.4%	0	0	0	0	0	0	0	0	0	0
177				Total Matches	3	5	5	5	4	5	5	4	4	5

Branan, J. (2023). A similar-sounding surname sequel: haplotype R-FT70038.
The Journal of the Fitzpatrick Clan Society 4, 11-44. doi:10.48151/fitzpatrickclansociety01323, Ver1, 9 June 2023