DNA Testing 101

An Introduction to the Various Tests Available and their Application to Genealogy

Presented by

Bob Mallett

Ottawa TMG Users Group

ottawa-tmg-ug.ca

Saturday, January 3, 2015

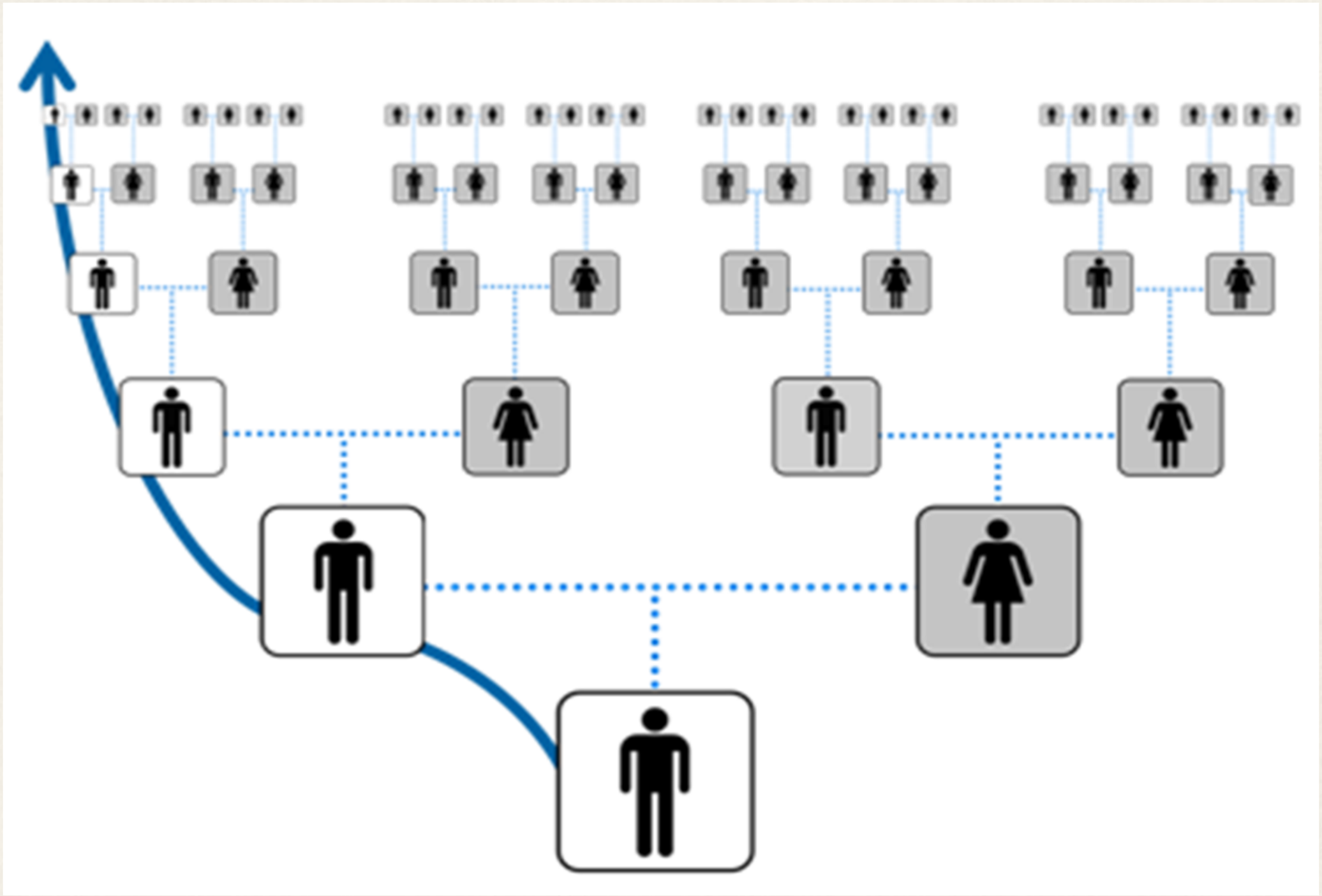
Introduction

There are 4 main categories of tests, each with its own application:

1. Y-STR — Male line only, father to son
2. Y-SNP or Deep Clade — Male line only, but (typically) deeper ancestry
3. Mithochondrial (mtDNA) — Female line only mother to child (son or daughter)
4. Autosomal — All lines

Y-STR

Y-Adam



Saturday, January 3, 2015

Y-STR

STRs (Short Tandem Repeats) are short sequences of DNA that are repeated numerous times at a given location on a particular chromosome. Mutations occur from time to time and the number of repeats can increase or decrease.

We simply count the number of repeats at a given location (a marker).

The actual count has no particular significance. What's important is how the counts compare to another's result.

Several levels of the test are available, ranging from 12 to 111 markers . The more markers the better, providing a more refined result. 37 markers is the norm now for an entry level test.

Estimates of relatedness between 2 individuals are based on the number of mismatches (genetic distance), and the mutation rates of markers, using a complicated algorithm.

The bottom line is that there can't be many mismatches.

Y-STR

Results from the Mallett Surname Y-DNA Study:

	DYS426	DYS455	DYS454	DYS388	DYS392	DYS438	DYS393	DYS437	YCAIIa	YCAIIb	DYS459a	DYS459b	DYS448	DYS394 (19)	DYS389 i	Y-GATA-H4	DYS385a	DYS385b	DYS389 ii	DYS447	DYS391	DYS390	DYS442	DYS460	DYS607	DYS439	DYS464a	DYS464b	DYS464c	DYS464 d	DYS456	DYS570	DYS458	DYS449	DYS576	DYS724 (CDYa)	DYS724 (CDYb)	ABS	FTDNA
Modal (South Tawton 1)	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	14	30	25	11	24	12	11	15	11	15	15	16	16	15	19	17	29	17	35	38		
South Tawton 2	12			12	13		13							14	13		12	14	30		11	24				11													0	0
South Tawton 3	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	13	14	30	25	11	24	12	11	15	11	15	15	16	16	15	19	17	29	18	35	38	2	2
Shebbear 1	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	14	30	25	11	24	12	11	15	11	15	15	16	16	16	19	17	29	17	35	37	2	2
Shebbear 2	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	14	30	25	11	24	12	11	15	11	15	15	16	16	15	19	17	29	17	35	37	1	1
Dowland	12	11	11	12	13		13	15			9	9	19	14	13		12	14	30	25	11	24				11	14	15	16	16			17	29				1	1
Maker	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	13	30	25	11	24	12	11	15	11	15	15	15	16	15	19	17	29	17	35	38	2	1
Mallet of Ash	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	14	30	25	10	24	12	12	15	11	15	15	16	16	15	19	17	29	17	35	38	2	2
Malet of St Audries	12	11	11	12	13	12	13	15	19	23	9	9	19	14	13	10	12	14	30	25	11	24	12	12	14	11	15	15	15	15	15	19	18	29	17	35	35	8	6
Alwington	12	11	11	12	13	12	13	15	19	23	9	10	19	14	14	12	10	14	30	25	11	24	12	10	15	12	15	17	17	17	16	17	16	29	17	38	40	21	17

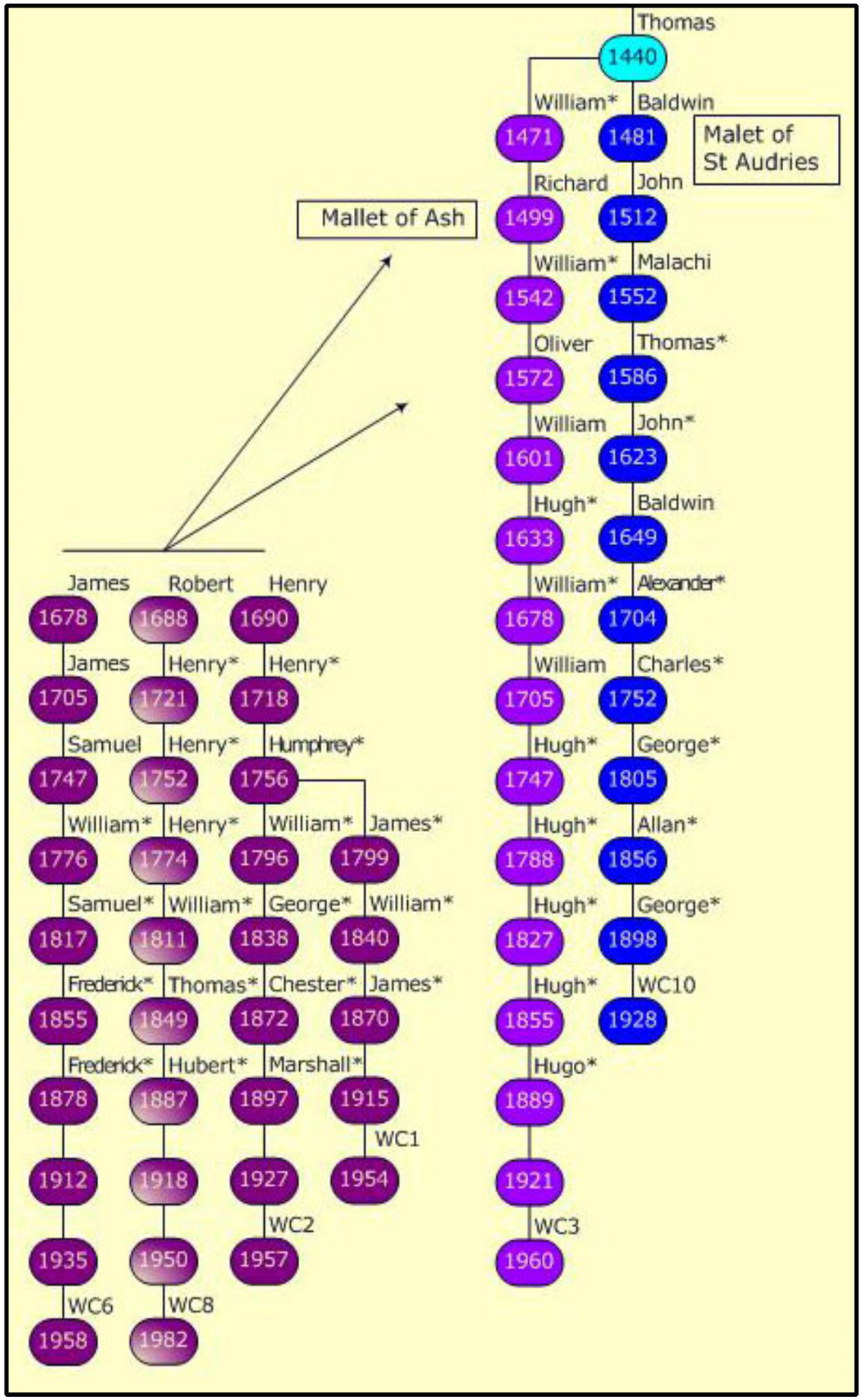
Probability of a relationship:

Genetic Distance = 1:

- 4 generations is 58.69%
- 8 generations is 88.85%
- 12 generations is 97.40%
- 16 generations is 99.44%
- 20 generations is 99.88%
- 24 generations is 99.98%

Genetic Distance = 6:

- 4 generations is 0.21%
- 8 generations is 5.17%
- 12 generations is 21.88%
- 16 generations is 46.07%
- 20 generations is 68.25%
- 24 generations is 83.57%



Devon / Cornwall and Mallet of Ash

The accompanying graphic illustrates the probable connection between three North Devon lines and the Mallet of Ash line suggested by the conventional research and bolstered by the DNA results.

The average genetic distance of the 3 families on the left compared to Mallet of Ash is 3, giving the following probabilities of relatedness:

...4 generations is 22.18%.

...8 generations is 66.49%.

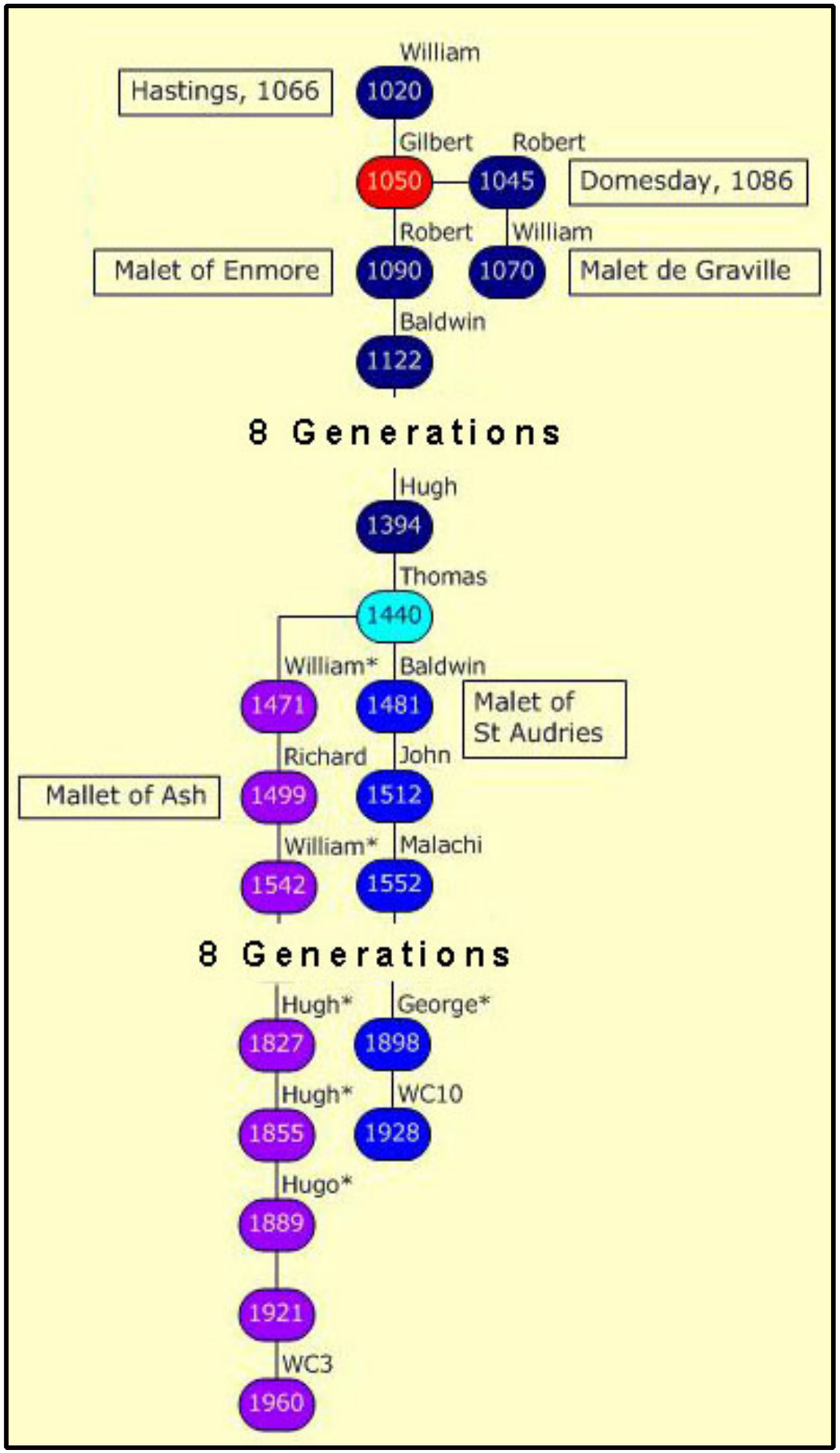
...12 generations is 89.97%.

...16 generations is 97.56%.

...20 generations is 99.48%.

...24 generations is 99.90%.

Recently a family of Malletts born in the mid 1600's was found in Bratton Clovelly, a village not far from the others, that provides a hard link for at least one of the families, and may be the missing link for all of them.



Malet of Enmore and St Audries (Somerset, England) Mallet of Ash (Devon, England)

The Malets first appear in Somerset in the mid 12th century. It is assumed that they are related to the Domesday Malets, but this has never been proven beyond doubt.

The first Malet in Somerset in connection with this line was Robert c 1090 at Enmore, in Somerset.

The Enmore line died out, but two spinoff branches continued; one was located at St Audries (West Quantoxhead) Somerset, the other in Iddesleigh, North Devon (Mallet of Ash). Both families are extant in the male line. Their probability of relatedness is as follows (GD of 6):

...4 generations is 0.89%.

...8 generations is 14.77%.

...12 generations is 45.23%.

...16 generations is 73.27%.

...20 generations is 89.43%.

Y-STR

Several other Mallett families have been identified that are genetically related to one another, but not to the other family groups, not even within the last 1000 years.

Some are tantalisingly close, but others are so far apart that there is absolutely no possibility of a relationship.

Many individuals have tested who match no-one else in the study.

There may have been a genetic relationship between at least some of the families at one time, but the link may have been broken because of an "NPE" (Non Paternal Event) or "Not the Parent Expected" ☺.

Conventional research might identify such an event, or it may be so far back in time that we will never know.

Saturday, January 3, 2015

Y-SNP

SNP = Single Nucleotide Polymorphism:

“Variation in the nucleotide allele at a certain nucleotide position in the human genome. When the change occurs it is called a polymorphism, and polymorphisms accumulate over time. A polymorphism can be very common (found in a significant fraction of global or localized populations) or very rare (found in a single individual). Common variations are used to track the evolution of the human genome over time (population genetics) and can be graphically represented in a haplogroup or phylogenetic tree.”

(From ISOGG Y-DNA Glossary).

ISOGG = International Society of Genetic Genealogy

Y-SNP

Unlike STRs, which can mutate at any time, once a SNP occurs, it remains — hence its value for population genetics and for genealogy.

Thousands of SNPs have been identified, and more are being discovered all the time as more people take the “Big Y” test at FTDNA , or “Geno 2.0” at National Geographic (there are other tests too).

Once a SNP has been identified it is placed on the Y-DNA Haplogroup tree at ISOGG and FTDNA (ISOGG is the most up to date). The tree begins with “Y-Chromosome Adam”, who is thought to have lived about 60,000 years ago. R-M269, the dominant haplogroup in Western Europe, is thought to have occurred 4000 to 10,000 years ago (Wikipedia: Haplogroup R1b).

Since the advent of “Big Y” and other tests, many downstream SNPs have been discovered, and it is thought that at least some of them will fall within a genealogical timeframe, making SNP testing useful to Genealogists as well as Population Geneticists.

Y-SNP

Your Predicted Haplogroup is R-M269

Search

The R-M269 lineage likely began in West Asia. It is the descendant of the major R-M343 lineage. Some of your genetic cousins traveled northwest into Central Asia and on to Europe. Others moved south and entered the Levant Region.

A B C D E F G H I J K L M N O P Q **R** S T

Tested Positive | Tested Negative | Presumed Positive | **Test Available** | Presumed Negative | Test in Progress

CTS5768 R-CTS5768
P25 L278 M343 M415 L774 L506 L754 L761 More... R-P25
L752 L502 L585 L389 P297 More... R-P297
M269 L483 L407 L500 L478 L773 L482 L265 More... R-M269
L23 L150 R-L23 Add
YSC0000072 R-YSC0000072 Add
L51 R-L51 Add
L11 L52 L151 P310 More... R-L151 Add
P311 R-P311 Add
P312 R-P312 Add
L21 More... R-L21 Add
DF21 R-DF21 Add
DF25 R-DF25 Add
DF5 More... R-DF5 Add
CTS3655 R-CTS3655 Add
L627 R-L627 Add
L626 R-L626 Add
L625 R-L625 Add
CTS1970 More... R-CTS1970 Add
PF4252 R-PF4252 Add
L130 R-L130 Add
DF23 R-DF23 Add
M222 BY196 Y2601 Y2602 FGC8738 R-M222 Add
FGC4077 R-FGC4077 Add
A725 R-A725 Add
FGC12948 R-FGC12948 Add
M4491 R-M4491 Add
A359 R-A359 Add

Y-SNP

Individual SNPs can be tested for about $30.00. But how does one know which SNP to test?

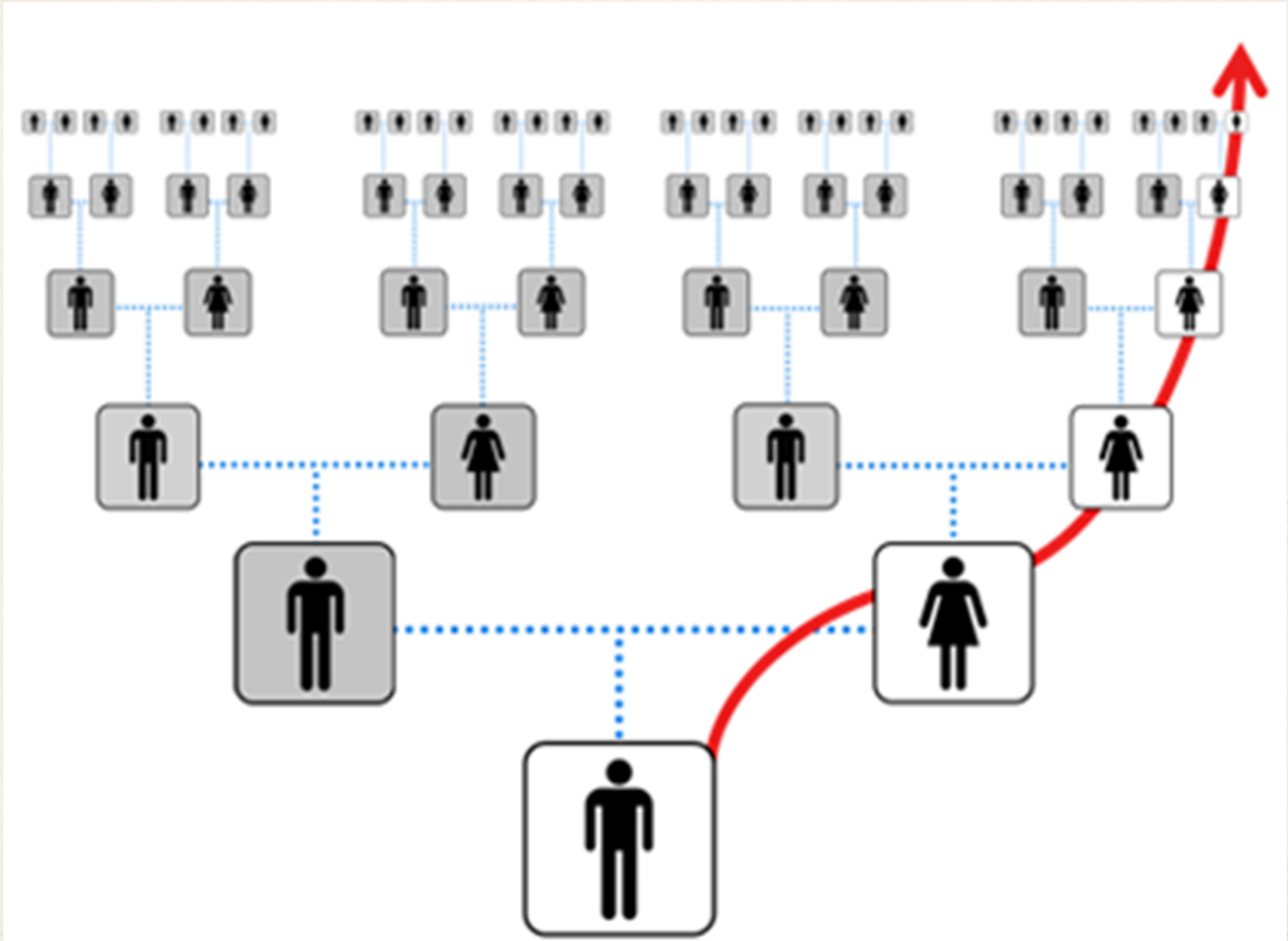
There exists a community of people trying to make SNP predictions based on STR values, and they are working with FTDNA to group likely SNPs together into “panels”, much like the 1 to 12, or 23 to 15 marker panels for STR testing, so that one can pursue SNP testing at a reasonable cost.

We’re not there yet. At the moment it’s the proverbial needle in a haystack.

Saturday, January 3, 2015

Mithochondrial

Mithochondrial Eve



Mithochondrial (mtDNA)

FTDNA currently offers 3 levels of mtDNA testing:

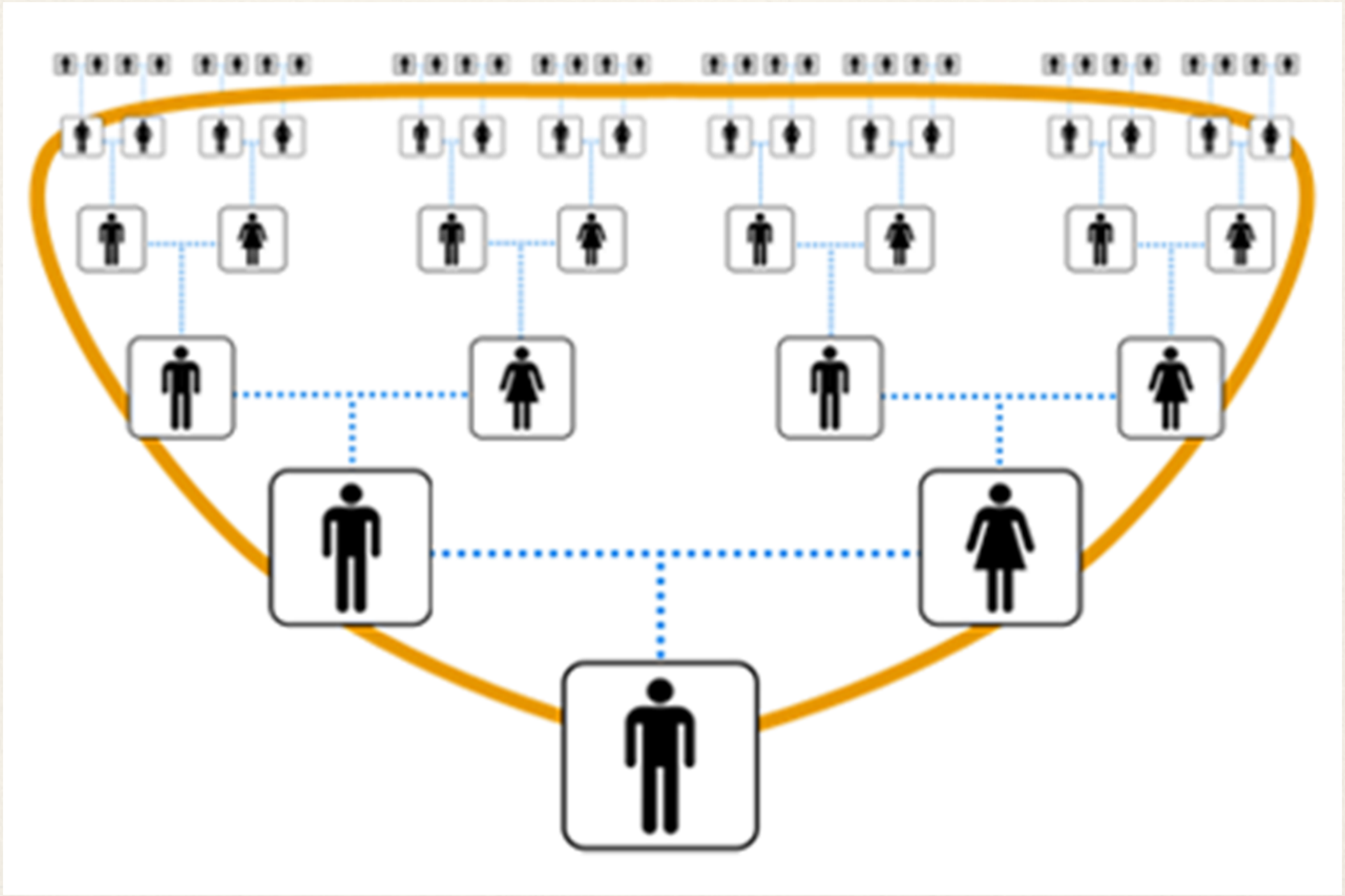
1. mtDNA
2. mtDNAPlus
3. mtFullSequence

Only mtFullSequence offers a probability of relationship at a high enough level to be really useful within genealogical time (95% confidence at 22 generations).

The other 2 tests can still be helpful as an adjunct to paper research, but they tend to point to deep ancestry. mtDNAPlus only allows for a confidence level of 50% at 28 generations. mtDNA only reaches 50% at 52 generations.

An mtDNA test can be interesting to determine one's deep maternal ancestry, e.g. from which of "The Seven Daughters of Eve" does one descend?

Autosomal



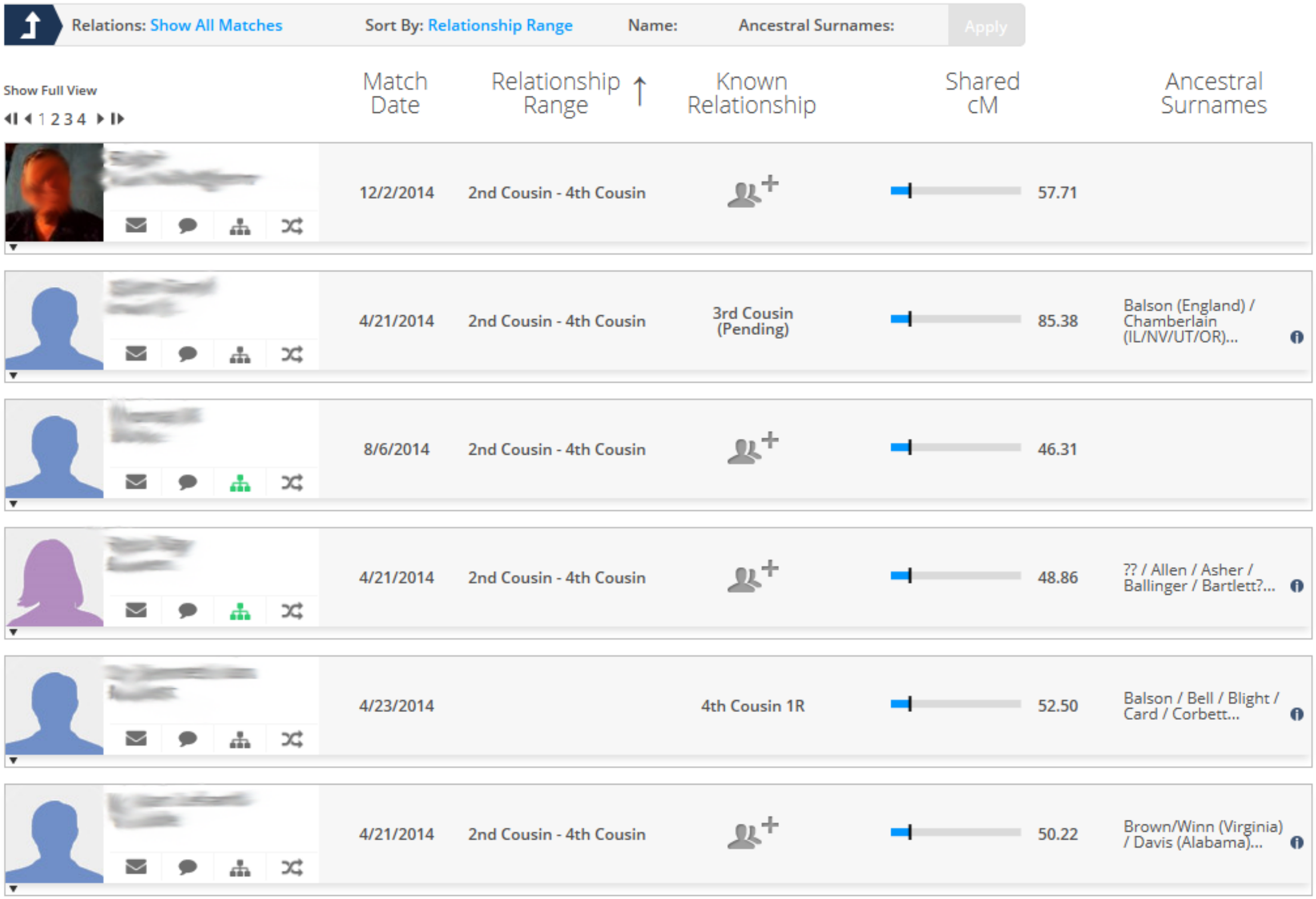
Saturday, January 3, 2015

Autosomal

Autosomal DNA statistics
50% Mother, father, siblings
25% Grandparents, aunts, uncles, half-siblings,double first cousins
12.5% First cousins
3.125% Second cousins
0.78% Third cousins
0.195% Fourth cousin
0.049% Fifth cousins
0.0122% Sixth cousins
0.003052% Seventh cousins
Chances of matching
2nd cousin >99%
3rd cousin > 90%
4th cousin > 50%
5th cousin > 10%
6th cousin or more distant Remote (typically less than 2%)

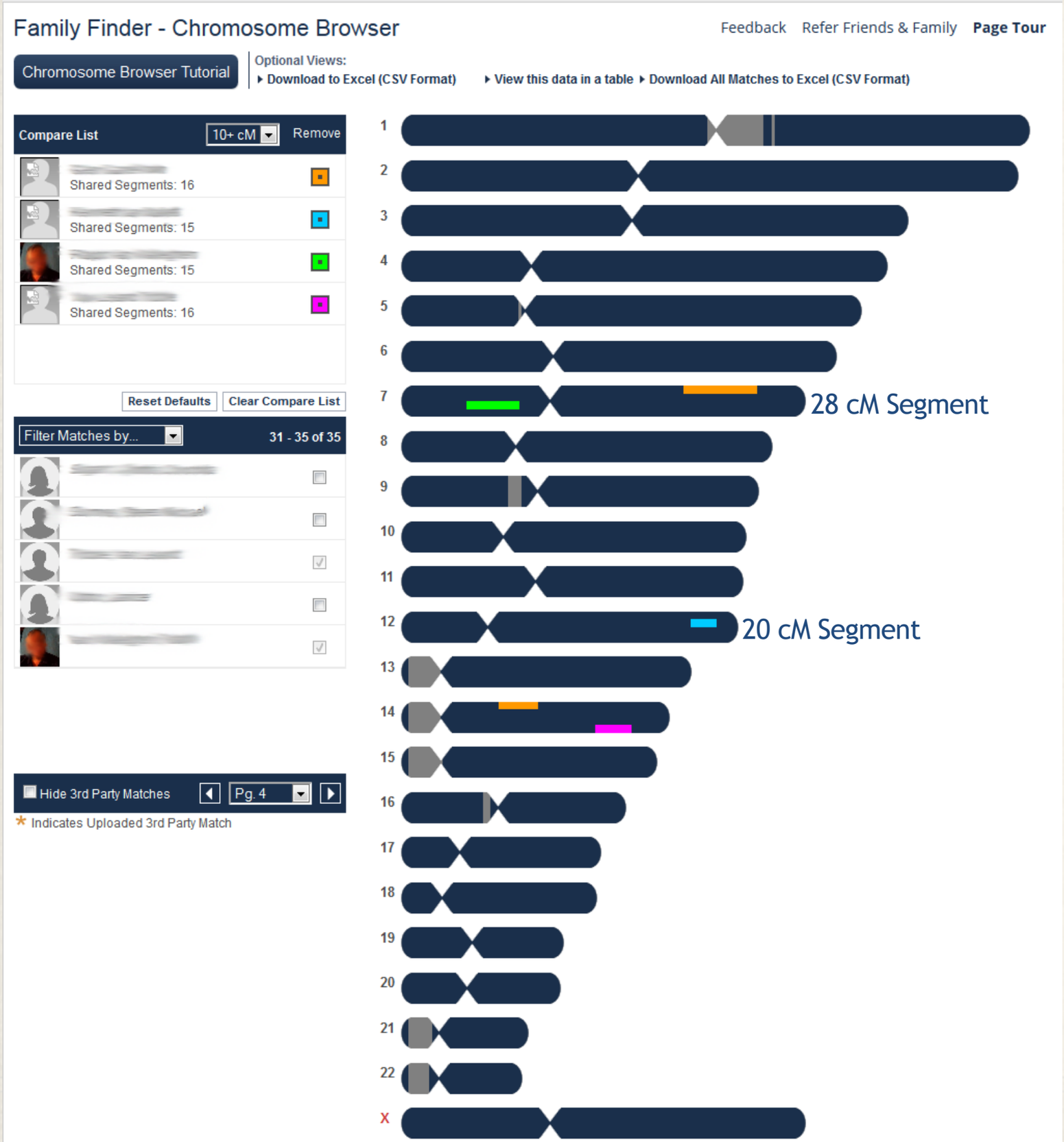
Matching segements are measured in Centimorgans (cM).

Autosomal



cM = Centimorgan: the distance between chromosome positions.

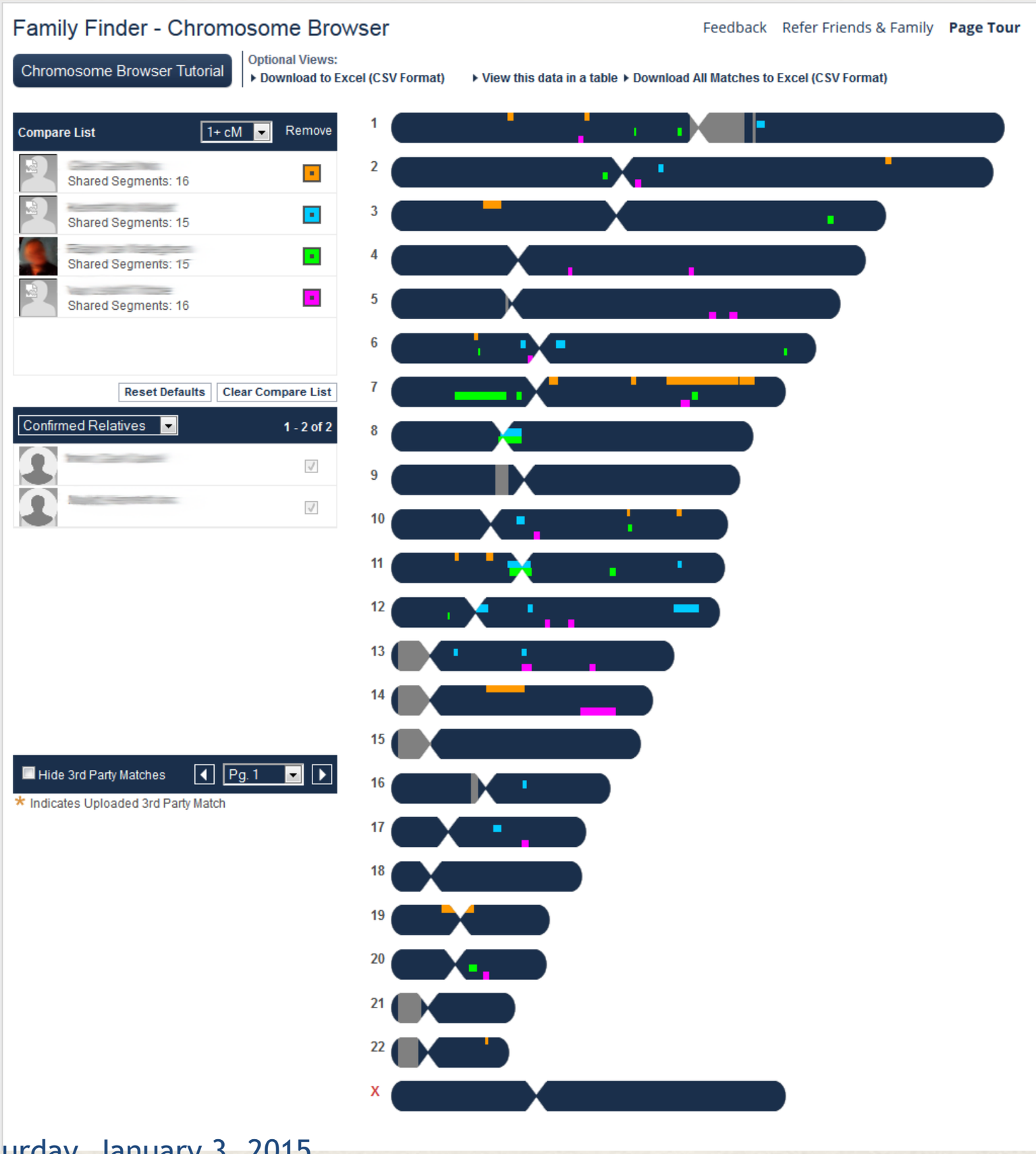
Autosomal



Filter: 10 cM
Prediction: 2nd to 4th Cousin

Actual Relationship?
Orange: 3rd Cousin
Blue: 4th Cousin once Removed
Green: No Idea
Pink: No Clue

Autosomal



Filter: 1 cM

Prediction: 2nd to 4th Cousin

Actual Relationship?

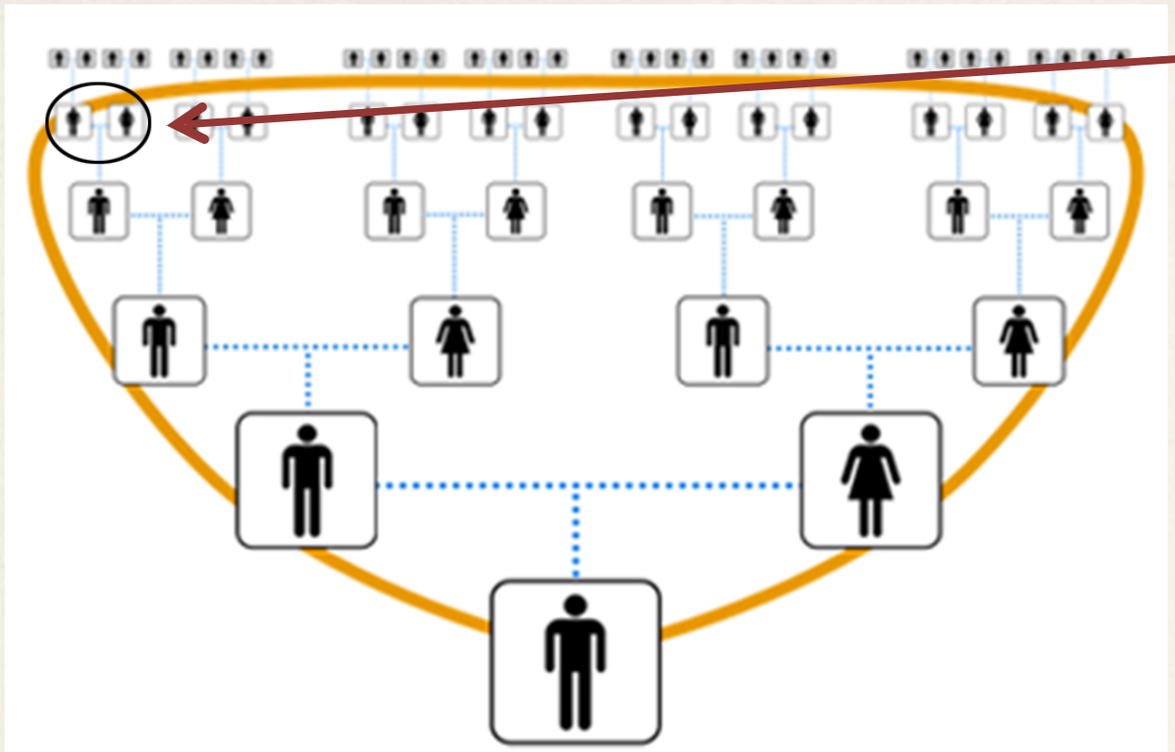
Orange: 3rd Cousin

Blue: 4th Cousin once Removed

Green: No Idea

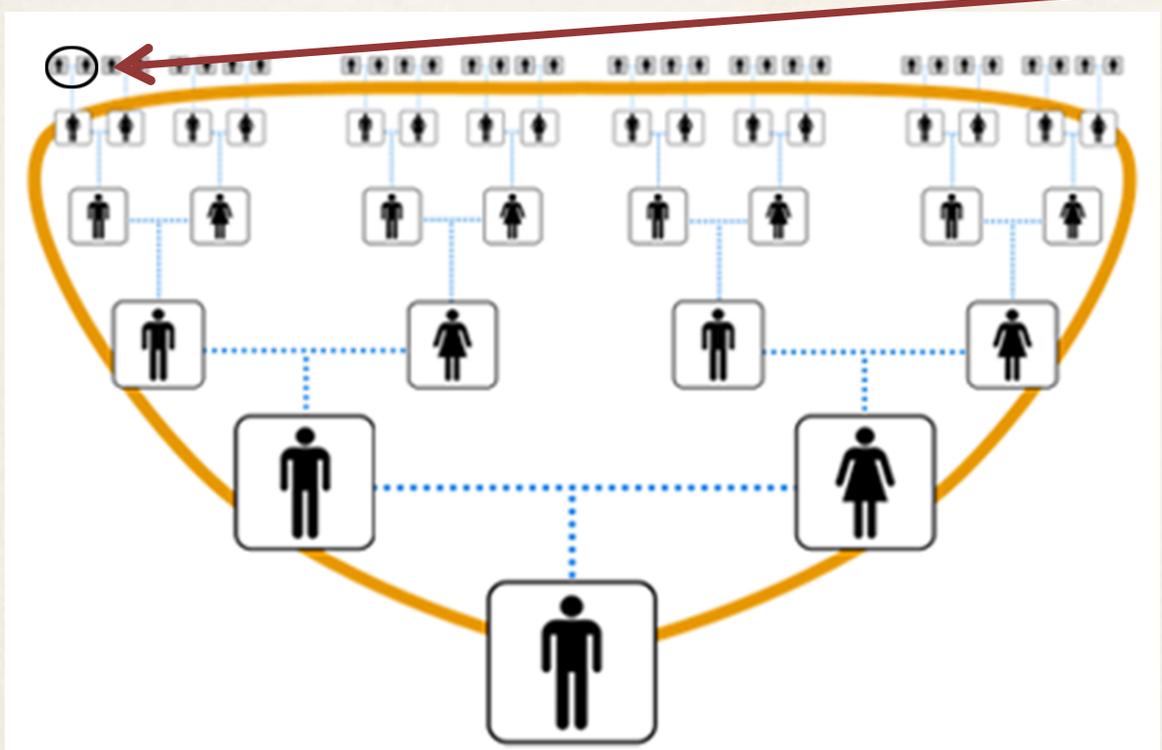
Pink: No Clue

Autosomal



3rd Cousin

Common ancestors:
GGGrandparents, 2 of 16 people



4th Cousin

Common ancestors:
GGGGrandparents, 2 of 32 people

Autosomal

What about the “Unknown” 2nd to 4th Cousins? Do we have a common ancestor?

The matching DNA could have come from any of the 56 people 3 generations and more back from me. The task of finding the common ancestor among so many possibilities is daunting. How does one go about it?

Research as many lines as possible, as far back as possible.

Look for shared surnames, and particularly shared rare surnames.

Look for shared geographical locations.

Concentrate on matches with the closest predicted relationships (fourth cousins and closer).

Focus on matches with shared segments of 10 cMs or above (20 cMs for Ashkenazi Jews).

Autosomal

The higher threshold for Ashkenazi Jews comes from the fact that they tended to marry within a restricted group. The same thing applies to families from remote areas, such as certain Newfoundland Outports, where a few families intermarried over many generations.

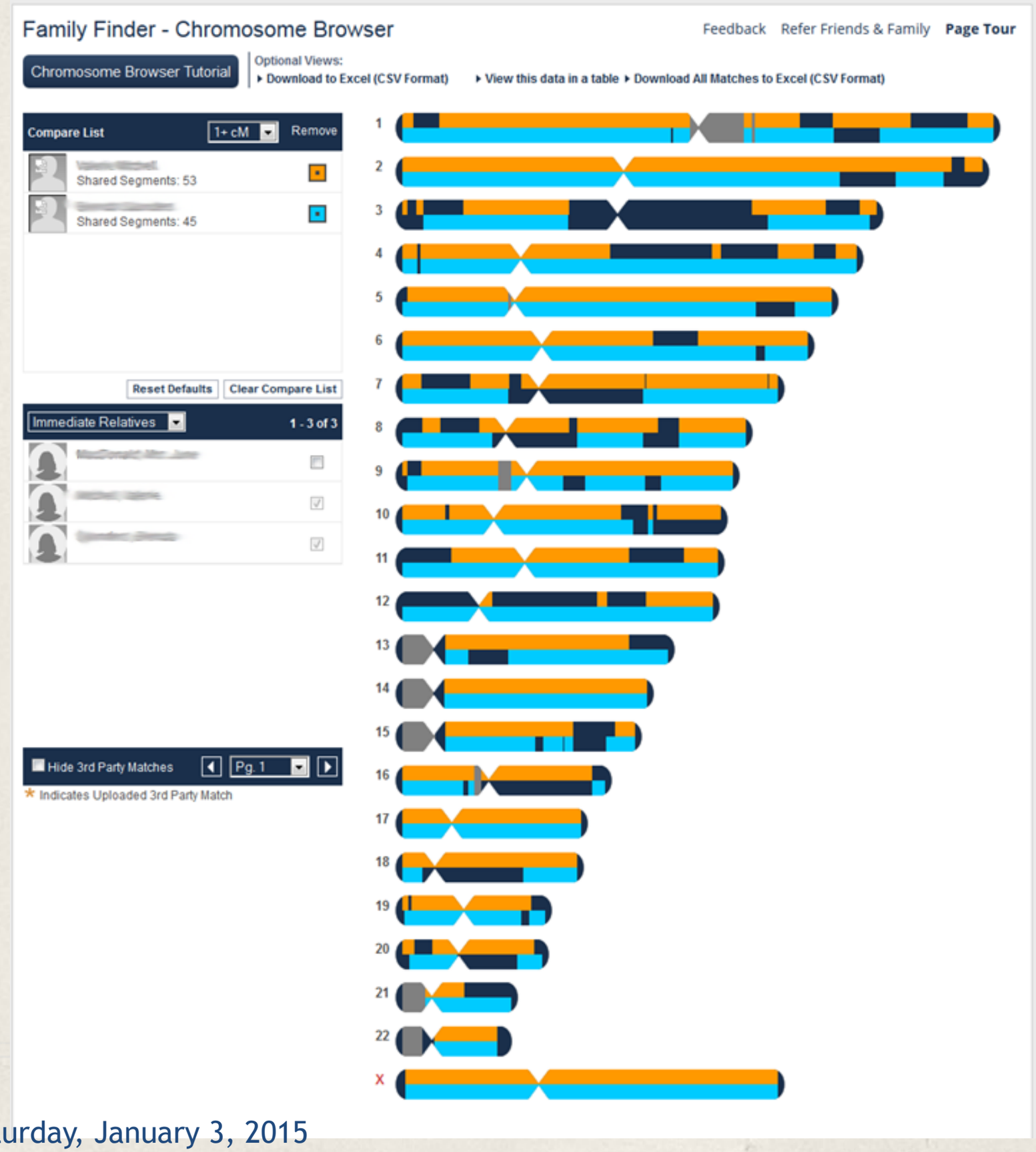
This sort of thing can lead to false positive matches, where segments are considered to be “Identical by State (IBS)”, not coming from a single common ancestor.

The term “Identical by Descent (IBD)” is used to describe matching segments that probably do come from a common ancestor.

There are many advanced techniques being developed and used to find common ancestors, too many and too complicated to go into here, but they all require a good deal of time, effort, and additional testing.

Tests are available from 23andMe, FTDNA, and AncestryDNA.

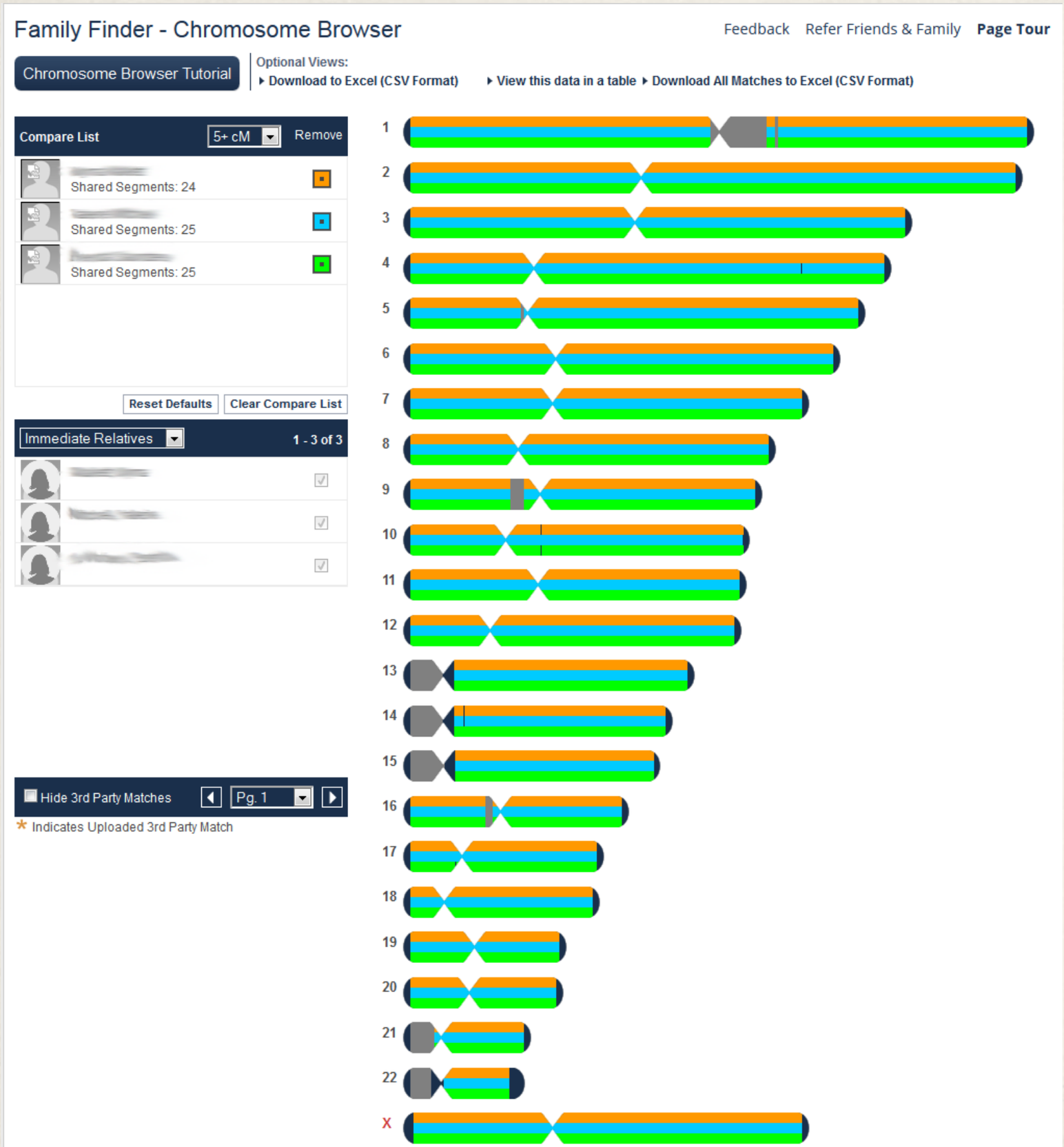
Autosomal



Filter: 1 cM
Prediction: Full Siblings

Actual Relationship?
Orange: Sister
Blue: Sister

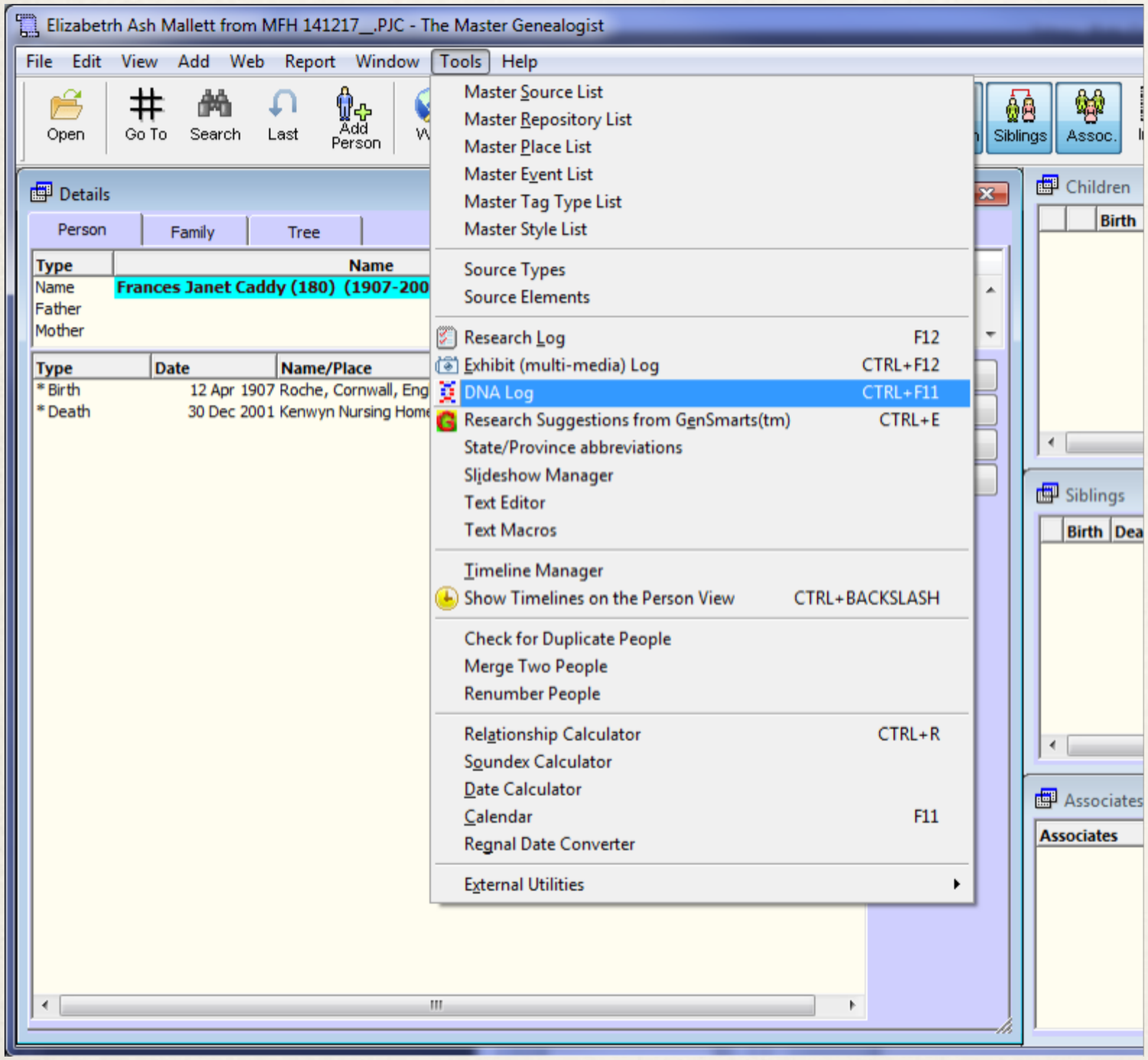
Autosomal



Filter: 1 cM
Prediction: Parent/Child

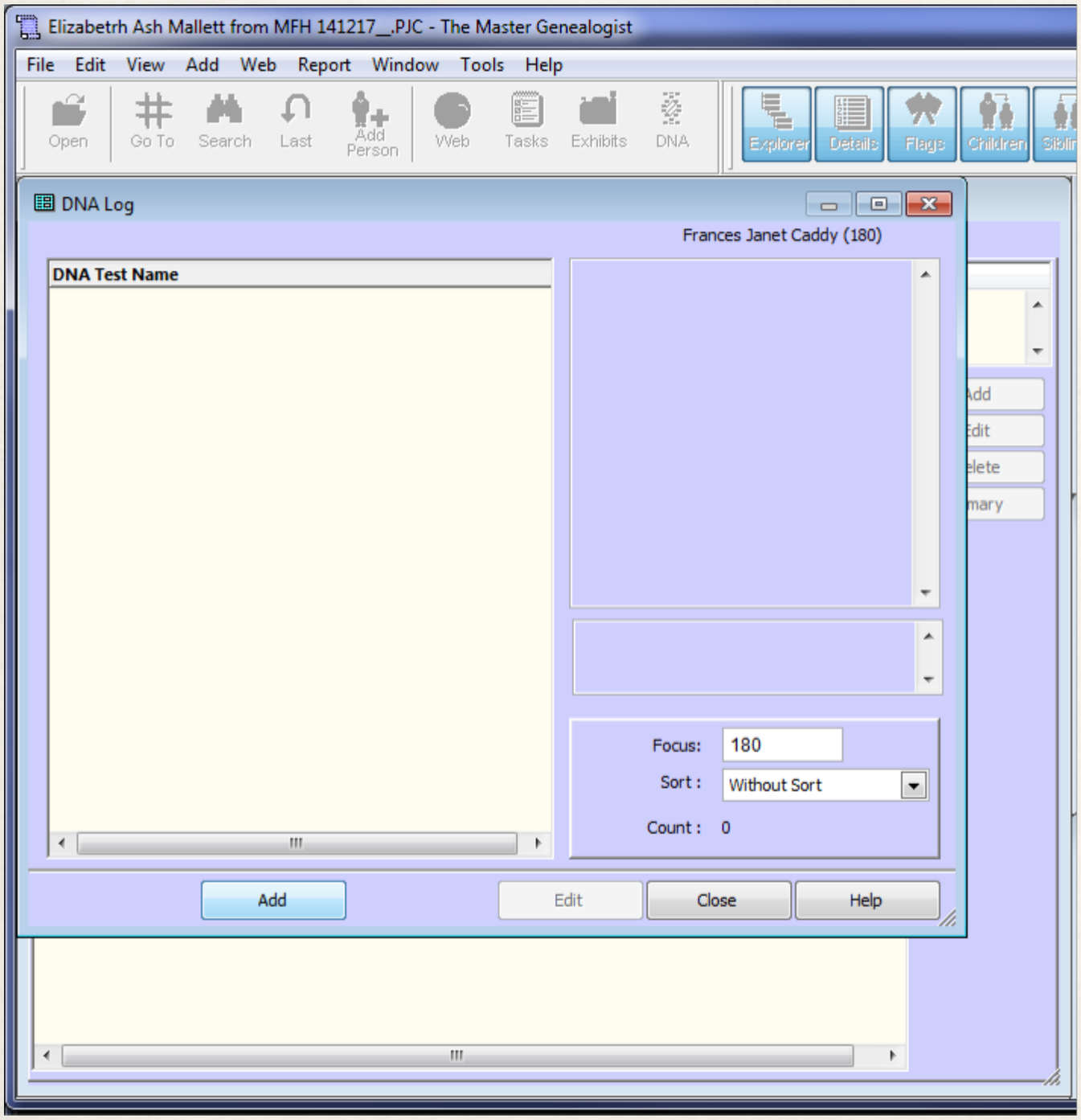
Actual Relationship?
Orange: Daughter
Blue: Daughter
Green: Daughter

TMG DNA Log



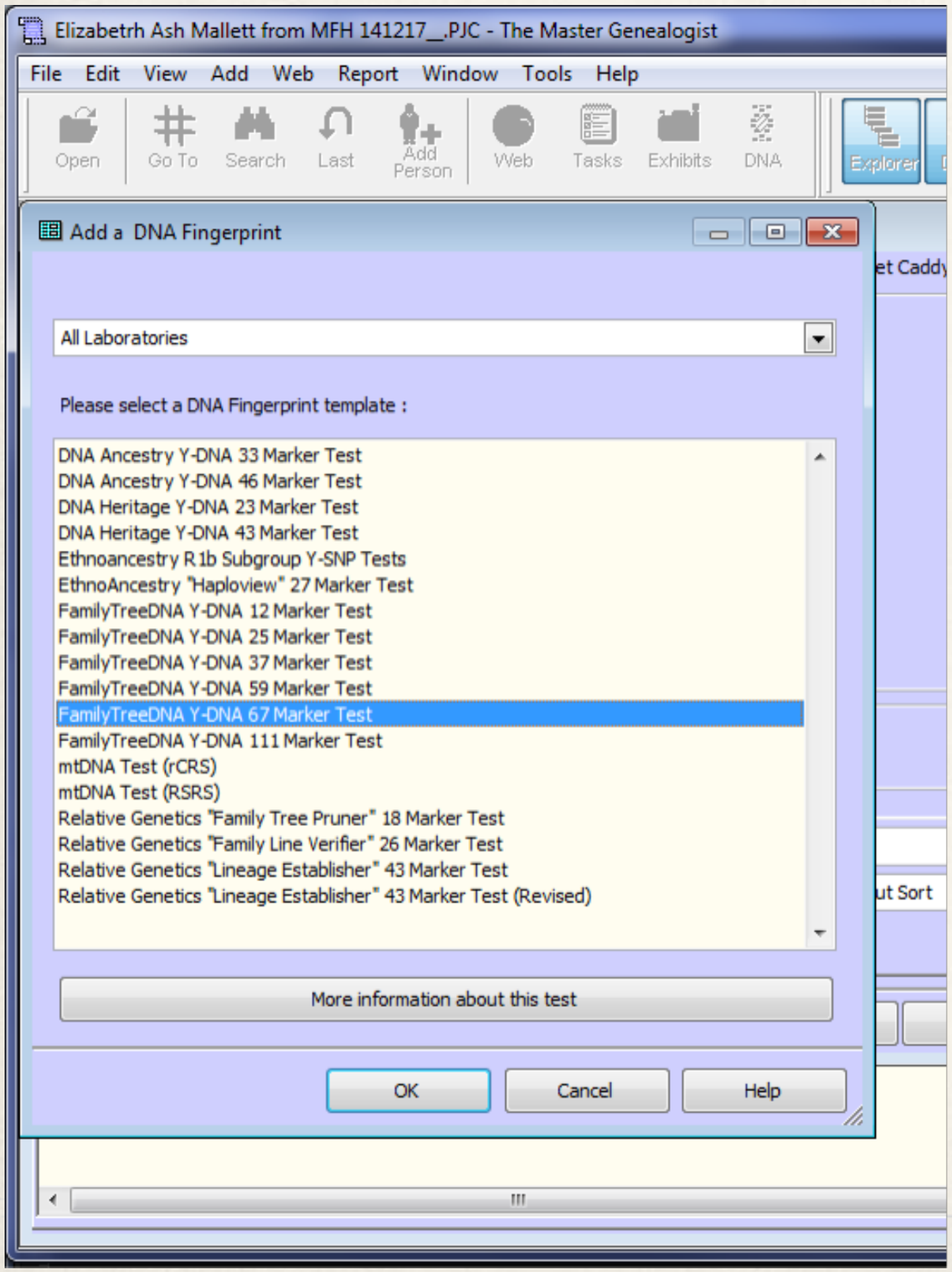
Go to the person you want to attach the log to, then choose Tools, DNA Log

TMG DNA Log



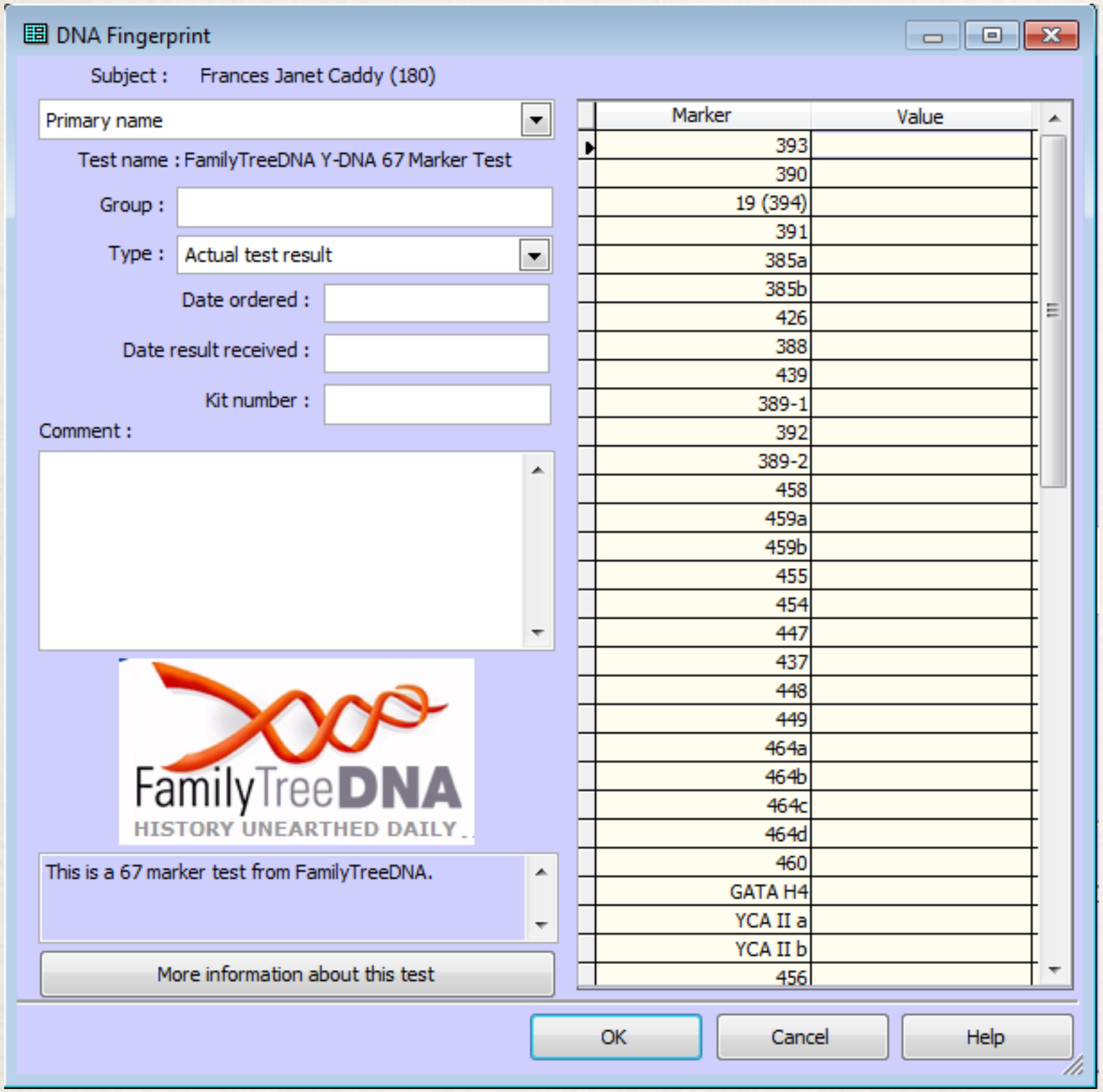
Click Add

TMG DNA Log



Choose the type of test.

TMG DNA Log



Enter the actual values from the test.

This is simply a log of results, attached to a particular individual, but nothing shows up in the person view.

One can enter a "group", but this just seems to be text. Different individuals are not linked by this value.