

Notes on the Cheshire Group STR Results

In 2014 I performed some Time to Most Recent Common Ancestor (TMRCA) calculations comparing the (then) 10 members of the Cheshire Group with each other, and with some closely matching Duttons. The results of these calculations can be found in four documents on the Warburton website:

- **Cheshire Group TMRCA Calculations**
- **Mutations Table**
- **Warburton-Dutton TMRCA Calculations**
- **TMRCA Calculations from 5 Dutton and Warburton 111 marker STR Results**

Before discussing the actual results, I have included a section on the method of conducting TMRCA calculations.

TMRCA Calculations

When calculating the Time to the Most Recent Common Ancestor (TMRCA) based on STR tests the factors involved include genetic distance, mutation rates, and the length of a generation. The criteria I used to perform these calculations are discussed below. For dates based on SNPs I have used the dates calculated by the R-U106 Project.

Mutation Rates

Mutation rates seem to vary considerably between markers, and maybe for a single marker between families. I revisited this question a number of times and put a document called **Mutations Table** on the Warburton Website with the details and sources for some of the published rates.

The mutation rates I used are based on those published by Marko Heinila, mainly because they were fairly recent (May 2012) when I did the calculations. Also they cover all 111 markers in the Family Tree DNA Y-DNA111 test. For each matchable set of markers I have calculated an average mutation rate from Heinila's individual marker rates as follows:

- FTDNA - 37 Markers - 0.36% - This seems quite high compared with the other rates but the 5 unique FTDNA markers are quite volatile.
- DNAH - 43 Markers - 0.25%
- Common - 32/34 Markers - 0.24%
- FTDNA 25 - 0.25%
- FTDNA 67 - 0.26%
- FTDNA 111 - 0.26%

The **Mutations Table** includes a table showing the the results of calculating the number of generations to the most recent common ancestor for each combination of genetic distance and number of markers compared when using average mutation rates based on Heinila.

Calculations

I have used Dean McGee's calculator at :

<http://www.mymcgee.com/tools/yutility111.html#Instructions> to perform the calculations, but not for calculating average mutation rates as it doesn't handle each element of multiple markers separately.

The Cheshire Group

The McGee calculator uses the Infinite Allele Mutation model to perform calculations. One drawback is that this method does not handle back mutations, so some people add an allowance (10% per 1000 years) to cover this. I have chosen to ignore this at this stage.

Also you can only choose to handle a difference of 2 or more between markers as either a single multi-step mutation, or a series of single step mutations. As in practice it is possible to distinguish some of these by comparing with the mode I have 'forced' the most likely genetic distance where necessary. I have counted two mutations where both participants differ from the mode, but treating a difference of 2 or more from the mode as a single mutation.

There is an alternative calculation method, Ken Nordtved's variance method, but this does not handle multi-step mutations and so tends to overestimate the TMRCA. Also I have yet to find a web-based calculator that uses this method.

As a cross-check I have also used Ann Turner's calculator (a Windows only downloadable .exe file) to calculate the likely number of mutations from the mode over the estimated 20 generations since the first Warburton was recorded, given the number of markers tested and the average mutation rate. These results are also in the **Mutations Table**.

Range

TMRCA calculations produce a range of probabilities. These ranges are quite wide. I have chosen to show a range which encompasses 1 standard deviation from the mean. There is a two thirds probability of any actual TMRCA falling within this range (i.e there is a 16% chance it falls before the earlier date, and an 84% chance it falls before the later date). On the other hand the R-U106 Project dates I have used for SNPs use a 90% probability range.

Date Calculation

A date can be obtained from the number of generations calculated above using the formula $SD - (YD * \text{Number of Generations})$.

SD is a start date based on the time of the last Y-chromosome transmission. 1947 is a commonly used date, but I have chosen to use my own date of birth which is 1945.

YD is the years per generation. Many sources seem to suggest numbers like 25 years per generation. However these tend to consider lines of inheritance and so consider only time to the first born son. When I look at my own ancestry the average generation over the last 10 generations is 37 years. In one instance my 3x great grandfather was born when his father was 59, and he was only the 4th of 8 children of his father's second marriage. I suspect there is probably quite a variation between different families. Some years ago I attended a wedding where the groom's grandmother, and the bride's great grandmother were present, and the groom's grandmother was the elder of the two.

I have seen 30 years quoted in relation to DNA calculations, and recent studies in Europe (particularly Iceland) are suggesting 33-37 years. I am using 35 years in my current calculations. This means that since the first person (Sir Peter) adopted the Warburton name in the 13th century we have had an average of 20 generations ($20 * 35 = 700$ years before 1945). Therefore to consider whether 2 modern Warburtons have a common ancestor we need to look at the probability of a common ancestor in the last 20 generations.

Cheshire Group Calculations

All Cheshire Group results are compared to a modal set of values. This is the most common value within the Cheshire Group for each marker, and it probably represents the values held by our common ancestor. Wherever a result includes a marker value which differs from the mode (a mutation), the marker and its value is shown in the Cheshire Group haplotree on the appropriate leg. Where a mutation occurs in more than one result it implies there is probably a more recent common ancestor.

The modal set of values for the Cheshire Group is characterised as follows:

1. The 12 markers in **Family Tree DNA** panel 1 conform to the Western Atlantic Modal Haplotype. This haplotype is extremely common on Europe's Atlantic seaboard.
2. Markers where the Cheshire Group share an uncommon value (allele) include DYS464=15-15-16-16, DYS437=12, and DYS449=30.
3. The modal values of all the markers in the FTDNA 67 marker test are identical to the modal values for the related group of Duttons. Comparison with modal values for the whole of **The King's Cluster** suggests this is also true for the FTDNA 111 marker test.
4. The results of DYF395S1=16-16, and DYS557=15 are indicators for membership of **The King's Cluster**. DYF395S1 is a double marker.

Some Short Tandem Repeat (STR) markers can be quite volatile and so may change more than once in the time period in question. It is assumed, wherever possible, that mutations only occur once. This includes one result of DYS438=10 which is 2 different from the mode of 12, but is assumed to be a single 2-step mutation. However DYS444 must have mutated from 12 to 11 on two separate occasions because these profiles fall within two separate groups that are defined by a shared mutation. DYS444 seems extremely volatile, having also mutated to 13 on one occasion, making 3 mutations in 7 tests of the marker. Mutations which occurred in two separate places are shown in italics on the **Cheshire Group Haplotree**.

I compared each of the 10 Cheshire group results with each of the others, giving 45 unique pairs. In some cases genealogical evidence exists to identify the most recent common ancestor, with greater or lesser confidence. There are also 2 groups that have a shared mutation, indicating a relatively more recent common ancestor. There then remain 32 unbounded pairs whose most recent common ancestor could be Sir Piers de Werberton, or any of his ancestors, though genealogical knowledge may put a boundary on the latest possible date.

Cheshire Group DNA Calculations on the Warburton website shows details of all the unique pairs including their genetic distance, the results of TMRCA calculations, and any boundaries on those dates derived from genealogical knowledge.

It includes details of the criteria used in the calculations, calculations of the probability of a given number of mutations from the mode over 20 generations (the likely number since the first Warburton lived), and weighted average dates for the 32 unbounded pairs.

The results show that the most likely number of mutations from the mode for both 37 and 43 marker tests is 2, but there are reasonable probabilities for any number from 0 to 5. The average date for the most recent common ancestor of the 32 unbounded pairs is 1177AD with a probable range (within 1 standard deviation) of 837AD to 1643AD. The 43 marker comparisons produce slightly later dates, whilst the fewer 37 marker comparisons produce earlier dates.

Warburton-Dutton Calculations

In February 2013 I discovered that the Cheshire Group closely match a group of 5 Duttons who are included in the **Dutton DNA Project** on **Family Tree DNA**. The historical connection between the Warburtons and the Duttons is described in detail in **Origins and Statistics** which can be found on the **Papers** page of the **Warburton Website**. Briefly the earliest known Warburton family were previously Duttons descended from Odard de Dutton, a Norman knight who arrived in England with the Norman Conquest.

The **Warburton Website** has 3 documents that compare the Dutton and Warburton results in detail. **A Comparison of Matching Warburton and Dutton Results** shows the actual DNA results, and identifies those markers which are significant identifiers of the group. It includes various modes for comparison.

The 10 Warburton and 5 Dutton results produce 50 Warburton-Dutton pairs who share a common ancestor. **Warburton-Dutton TMRCA Calculations** is a document which shows the results of TMRCA calculations for each pair, and produces a weighted average date from the 26 pairs where at least 37 markers could be compared. This produced an average date of 1109AD, with a 68% (1 standard deviation) probability that he lived between 786AD and 1529AD. This compares with the historical evidence that he would have been born between circa 1040 and 1200. Details of the criteria used in the calculations are included in the document.

TMRCA Calculations from 5 Dutton and Warburton 111 marker STR Results is a document that looks specifically at 5 111-marker tests, 2 Dutton and 3 Warburton and produces a similar calculation, though the results are not averaged. The six Warburton-Dutton comparisons produce average dates in the range 965AD to 1210AD.

All of these results tally closely with the historical evidence which dates the Dutton-Warburton common ancestor between circa 1040 (birth of Odard de Dutton) and 1200 (birth of Geoffrey de Dutton, whose son Piers adopted the Warburton name). The dates in the document are slightly different.