

The HAWGOOD FAMILY DNA STUDY

The Hawgood family DNA study aims to connect lines of Hawgoods into the main family tree. The following document explains how this works and the results of the study.

www.hawgoodfamily.co.uk

Introduction

Where did we come from?

Early Origins

All humans living today have their earliest ancestors around 100,000 years ago in eastern Africa. Approximately 50,000 years ago they left Africa and migrated to Asia and beyond. Over time, a number of specific genetic groupings emerged defined as Haplotypes. Within each general Haplotype are subgroups and further subgroups which can narrow down specific relationships between surnames.

Hawgood DNA comes from Haplogroup I1-M170, which emerged in Europe about 28,000 years ago. Around 10,000 years later, the ice age had a major impact on our ancestors in Haplogroup I when most of northern and central Europe was uninhabitable. This forced them to retreat to refuge areas in Iberia and the Balkans where living conditions were better.

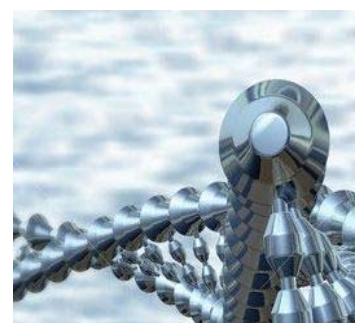
Post ice age

As the Ice Age receded, our ancestors of Haplogroup I1-M170 moved out into the surrounding areas. As Europe was being repopulated, some 8,000 years ago near Denmark, a sub group of I-M170 emerged, I1-M253, of which Hawgoods are a member of.

These ancestors migrated west into the area of the Doggerland land bridge, a piece of land which linked England to Northern Europe, including Denmark and north Germany, and is now covered by the North Sea. It was therefore easy to move into England, where written records of Hawgoods can be traced from around 1500.

Today, the subgroup of I1-M253 is relatively common in Scandinavia at around 35-40% of the population, as shown in the map above, increasing to up to 50% in certain Finnish provinces.

When analysing the specific incidence of the Hawgood values for markers within the overall 'I' population, one of the interesting outcomes is that whilst still being in I1-M253, Hawgood DNA has several marker values that are relatively uncommon. Hence finding similar persons that match the other markers, and also with this combination would be quite significant.



Part 1 - An explanation of DNA testing

What is DNA testing?

An overview

DNA testing can take lots of forms, but genealogical testing only looks at tiny 'junk' fillers in the DNA of the Y-chromosome. These areas are very useful to examine as they rarely change (or mutate) between generations. If these points, or what are known as 'markers', are the same in two males, they are related. The test is known as a Y-STR test, where STR stands for 'short tandem repeat' which is explained below. The test is conducted by sending a cheek swab back to the testing company.

How mutations occur

A mutation is simply a change in the DNA sequence which occurs when a cell is dividing and a certain enzyme fails to copy the DNA correctly. Cells divide through a process called mitosis, where the DNA makes a copy of itself and passes it to the new cell, using an enzyme called DNA Polymerase. This enzyme reads the original code along the chain, and builds the new strand of DNA. The code is made up of a very long pattern of four different nucleotides, abbreviated to A,T,C, and G.

At certain points, the DNA code repeats itself (this is a short tandem repeat, or STR) and sometimes there is a slippage so that for example instead of 7 repeats, we can get 8. This error does not happen very often which makes this kind of analysis very useful to determine relationships and ancestry. An error in these areas does not affect the operation of the cell as it is contained in a junk region of the DNA.

Test result format

The test results themselves are meaningless, except for comparison to another person. The naming convention for each marker is usually a code prefixed by 'DYS', and the test result for each marker is assigned a number equal to the number of times that the DNA sequence is repeated at that location. For example where $DYS455=7$, the marker point is named $DYS455$, and the value of 7 means that the DNA code is repeated 7 times. If the value of two people tested at location $DYS455$ is the same, there is a match. The number of matches, can be used to calculate the TMRCA - the length of time to most recent common ancestor.

The Hawgood family study uses genebase.com examining either 44 or 67 markers.

What do the results tell us?

If two people are closely related, all or nearly all of the markers will be the same. The further apart they are, the more differences, or mutations, will exist. As a rough rule of thumb, if 67 markers are tested, there would be one mutation every 100 or so years.

If we do not know the generation gap between two people, we can calculate the expected gap using the number of mutations and the average mutation rate.

Are you concerned about DNA testing?

Not forensic

Genealogical Y-STR DNA testing is NOT the same as forensic DNA testing undertaken in police investigations. Forensic tests look at genetic profiles that are not held on the Y chromosome and thus are not used in genealogical Y-STR DNA testing.

Not medical

Genealogical Y-STR DNA testing is NOT the same as medical DNA testing which aims at diagnosing genetic disorders looking at active gene portions of our chromosomes.

Not paternity

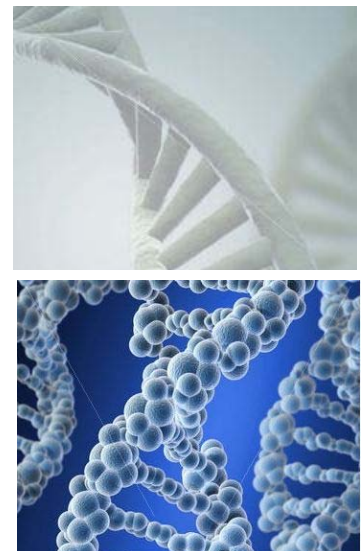
Genealogical Y-STR DNA testing is NOT the same as paternity DNA testing which spreads the test over several chromosomes and is not confined to the Y chromosome as with genealogical testing.

However, genealogical testing can reveal some unexpected male family relationship facts eg two brothers take the same genealogical DNA test, it could be shown that they don't have the same father. Nothing is revealed regarding any family relationships involving females.

Junk fillers

Genealogical Y-STR DNA testing examines tiny sections from the filler (junk) DNA of the Y chromosome which does not yield any direct information about the active genes of the Y chromosome.

Genealogical testing therefore is not used to show genetic disorders caused by abnormalities in genes on other chromosomes or on the Y chromosome.



Mutation rates

Speed is important

The speed of mutation, or what is known as the mutation rate, over which there is much confusion and divided opinion, is the key element in calculating how far apart two people are related.

Early studies (Walsh 2001) indicated an average mutation rate of 0.2% (meaning that in a 44 marker test, a marker would mutate once every 350 years) but later studies a rate of 0.4% (once every 170 years). Some more recent data indicates rates in excess of 0.5% (once every 140 years).

Accuracy is important

The problem with using a rate of 0.2% instead of 0.4% is that the common ancestor calculation will be twice as far apart. Some testing companies however still use the original Walsh value of 0.2% to predict generation gaps. If the input is flawed, then so will be the output. We need therefore to work out an accurate input to be confident of the calculated results.

Some of the difficulties arising are:

- Each marker has a different individual mutation rate, some fast, some slow
- Different testing companies include different markers in their tests, which means that the average rate of mutation from each company will differ. For example there are a larger number of fast mutating markers in the FTDNA 37 marker test, than the Genebase 44 marker test.
- Some studies have small sample sizes and this does not always provide reliable data

Overcoming the inconsistencies

There are tens of thousands of test results and more being added all the time. The good news is that studies are revealing mutation rates which are broadly consistent. A slow mutating marker may have a variance between different studies, but it will still be, broadly speaking a slow mutating marker. By the same token a fast mutating marker in one study does not suddenly become a slow mutating marker in another, it still remains a fast marker across different tests.

Combine multiple databases

We have taken results for individual markers from as many sources as possible, from wide private studies to published testing company data, and then for each individual marker taken the average of all sources. For each marker we have a specific mutation rate which is based on a massive amount of data (some 75-100,000 results). With such a large database, the calculation of the most common recent ancestor becomes much more reliable.

Current average mutation rates

The database is constantly being refined, but currently, the Genebase 44 marker test has an average mutation rate of approx 0.0027, or 0.27%, and the 67 marker test, approx 0.0041, or 0.41%. If you use a different company than Genebase, then you can still use the data on our website for each marker and using your company's specific markers, overlay this data and calculate an average mutation rate for your particular test.

Calculating how many mutations to expect

A simpler explanation

The concept of marker mutations confuses many at first glance. But it is relatively simple and can be explained using the technique of 'expected values', which is the long-run value taken over many independent repetitions.

A lottery

Consider a lottery draw that has numbered balls from 1 to 300. You have bet on number 50. When the draw is held you would expect your number would be unlikely to come up. Your odds are 1/300, or 0.33%

Say that in the lottery, not one, but 44 balls are drawn. The odds of your number coming up are now 44/300 or roughly 1/7. If you then do the same 44 draw every day for a week, you would expect that your number would come up statistically 7 times \times 1/7 odds = 1, or in other words you would expect to win once in that week.

Apply to STR markers

Now consider that the 44 balls are in fact the STR markers and that the 7 days are 7 generations. Using the same calculation, you would expect during the 7 generations that 1 mutation would occur on your own 44 markers.

When comparing two people, each one could have 1 mutation, so the total expected mutations must be double that for one person, which is 2 mutations. So comparing two people over 7 generations with 44 markers tested, the expected number of mutations is 2.

This can be distilled to a formula:

No of Markers tested x (No of generations x number of persons tested) x average mutation rate.

Sometimes the phrase 'No of generations x number of persons tested' is rewritten as 'No of transmission events'. The transmission event value is simply the number of generations \times the number of persons being compared. So for two 1st cousins, the most recent ancestor would be their grandfather. This would create 2 transmission events each, being grandfather to father to son for each cousin. The total transmission events would be 4.

This formula can be used to create a table of expected mutations. Using the table, for example, in a 63 marker test, using a mutation rate of 0.41%, the number of expected mutations where 11 generations exist between two parties tested, would be 6 (rounded as 5.7 mutations is not possible). The expected result would therefore be 57/63 matching.

Rate	0.27	0.41
Markers	43	63
Generations	Expected mutations	
1	0.2	0.5
2	0.5	1.0
3	0.7	1.6
4	0.9	2.1
5	1.2	2.6
6	1.4	3.1
7	1.6	3.6
8	1.9	4.2
9	2.1	4.7
10	2.3	5.2
11	2.6	5.7
12	2.8	6.2
13	3.0	6.8
14	3.3	7.3
15	3.5	7.8
16	3.8	8.3
17	4.0	8.9

More on the maths

Likelihood of expected outcome

The expected result may not necessarily be the actual result. Throwing six dice should produce one six, but it may not. By the same token, in our 44 marker test over 7 generations with an average mutation rate of 0.33%, we may not see the expected result of 2 mutations.

How likely is the expected result? We can calculate this by using the following formula :

$$(1 - P)^{(T-t)} \times P^t \times \frac{T!}{(T-t)! \times t!}$$

Where :

P = the mutation rate (quoted as 0.0033 and not a percentage at 0.33)

T = the total number of markers tested multiplied by the number of generations x the number of people in the comparison

t = the number of markers that have mutated

! = factorial (6! is the same as 6x5x4x3x2x1)

To work out the probability of 2 mutations in 44 markers over 7 generations, between 2 people using a mutation rate of 0.33%:

$$\begin{aligned} P &= 0.0033 \\ T &= 44 \times 7 \times 2 = 616 \\ t &= 2 \end{aligned}$$

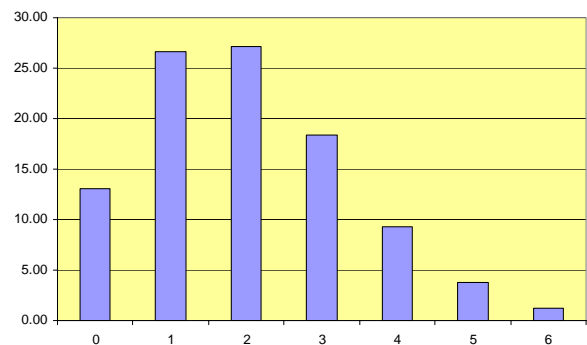
Plugging the numbers in to the equation

$$\begin{aligned} &= (1-0.0033)^{(616-2)} \times 0.0033^2 \\ &\quad \times (616!/614!)/2! \\ &= 0.0131394 \times 0.00001089 \\ &\quad \times (616 \times 615)/(2 \times 1) \\ &= 0.0271 \text{ or } 27.1\% \end{aligned}$$

Note that a calculator will not compute 616!, as the output is too big to handle. 616!/614! is the same as 616x615. Using another example, 64!/60! is the same as 64 x 63 x 62 x 61.

The table below shows the percentage chance of 0,1,2,3,4,5, and 6 mutations in a test of 44 markers where the two parties are 7 generations apart. This shows that whilst 2 mutations is the most likely outcome, 1 mutation would not be comparatively unlikely.

Percentage chance of number of mutations



Care with one set of results

Using a small dataset is not always reliable and in order for results to be meaningful, one must consider any one set of results in the context of other results.

The more participants, the more likely that the expected results statistically will pan out in reality. Expected maths always works in the long run but needs a significant dataset.

To avoid manual calculations see the website which provides a calculator for the above.

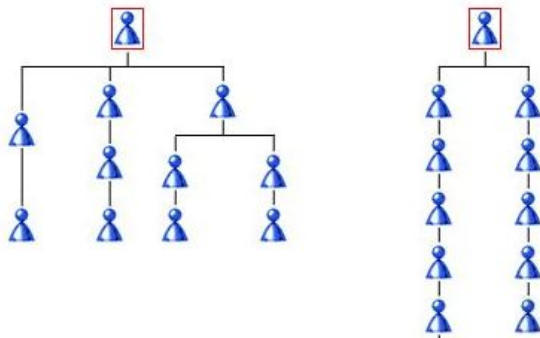
Caution on transmission events

Counting transmission events

There is not need to consider this when two people are being examined. The number of transmission events is simply the number of generations x 2.

Where there are more than two people being examined, the number of transmission events is not necessarily the number of generations multiplied by the number of people tested.

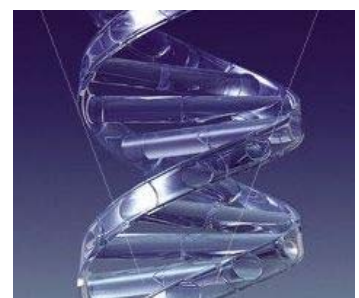
The diagram shows two examples where there are 10 transmission events.



The example on the right has the same common ancestor and the transmission events are simply 5×2 . However, if those being tested do not have the same earliest common ancestor, as in the left example, then the calculation is not as straight forward. In this example there are four people tested where the transmission events are not 12 (4×3) but are still 10.

This is as the third and fourth people share a common ancestor one generation later than the first and second.

Therefore to calculate the number of transmission events for more than two people, where they do not all share the same common ancestor, the events must be counted manually to ensure that there is no double counting.



Part 2 - The results of the Hawgood DNA study

On the next page is a list of numbers for the individuals tested. We are simply looking for values that match, and the number of mismatches is a guide, and a guide only, to the distance apart.

We are also looking for same mutations occurring in more than one person, as this usually represents a new branch. It is possibly but very unlikely that the same mutation will occur by chance in unrelated branches, except with the very fast mutating markers such as DYS724a and DYS724b whose mutation rates are up to 10 times faster than the 'normal' mutation speed.

Looking at the table on the next page, it can be seen that DYS19 is the same for all individuals, as are the next dozen or so.

DYS449 is the first variation and this is seen in Ma and Mi who are known third cousins. We therefore know that this mutation occurred in 1842 or earlier, as this is their first common ancestor.

The most significant variation is on DYS464 which is a polymorphic marker with at least four different values. The normal values for Hawgood DNA and indeed many of the same Haplotype (I1) are 12,14,15 and 16. However a mutation has occurred which is seen in four individuals which have different values, and within two of the four a slight variation there as well.

This shows that there are in fact **two distinct branches** and we can conclude that these branches are from two of the sons of John Hawgood born 1663.

C and G from the 12,14,15,16 group, and descended from William Hawgood born 1704. T and Ar are also likely to be from this line, but it is not certain yet their common ancestor is after or before 1704. The likely hood is that it is not earlier than 1704 as the number of mutations is too small.

D and 'An' are descended from Thomas born 1706, and the uniqueness of the mutations on DYS464 means we can be confident that Ma and Mi are also descended from Thomas.

Detailed results

	Marker test	Deduced Core DNA
DYS19	44	15
DYS385a	44	13
DYS385b	44	14
DYS388	44	14
DYS389i	44	13
DYS389ii	44	30
DYS390	44	22
DYS391	44	10
DYS392	44	11
DYS393	44	13
DYS426	44	11
DYS437	44	16
DYS438	44	10
DYS442	44	12
DYS445	44	11
DYS446	44	12
DYS447	44	23
DYS448	44	20
DYS449	44	29
DYS453	44	11
DYS454	44	11
DYS455	44	8
DYS456	44	14
DYS458	44	15
DYS459a	44	8
DYS459b	44	9
DYS460	44	11
DYS461	44	12
DYS462	44	12
DYS468	44	27
DYS484	44	13
DYS522	44	11
DYS527a	44	20
DYS527b	44	21
DYS531	44	11
DYS557	44	16
DYS588	44	19
GATAA10	44	13
GATAA4	44	12
GATAC4	44	22
GATAH4	44	11
YCAIIa	44	19
YCAIIb	44	21
DYS464a	44	12
DYS464b	67	14
DYS464c	67	15
DYS464d	67	16
DYS413a	-	23
DYS413b	-	25
DYS436	67	12
DYS444	67	13
DYS452	67	31
DYS463	67	21
DYS472	67	8
DYS481	67	25
DYS511	67	9
DYS518	67	21
DYS520	67	21
DYS537	67	11
DYS570	67	19
DYS576	67	18
DYS590	67	8
DYS607	67	13
DYS612	67	35
DYS614	67	29
DYS644	67	16
DYS710	67	14
DYS711	67	33
DYS724a	67	32
DYS724b	67	36

Ma	Mi	D	An	C	G	T	Ar	Haywood
15	15	15	15		15	15	15	15
13	13	13	13	13	13	13	13	13
14	14	14	14	14	14	14	14	14
14	14	14	14	14	14	14	14	14
13	13	13	13	13	13	13	13	13
30	30	30	30	30	30	30		30
22	22	22	22	22	22	22	22	22
10	10	10	10	10	10	10	10	10
11		11		11		11		11
13	13	13	13	13	13	13	13	13
11		11	11		11	11	11	11
16	16	16	16	16	16	16	16	16
10	10	10	10	10	10	10	10	10
12	12	12	12	12	12	12	12	12
11	11	11	11	11	11	11	11	11
12	12	12	12	12	12	12	12	12
23	23	23	23	23	23	23	23	23
20	20	20	20	20	20	20	20	20
30	30	29	28	29	29	29		29
11	11	11	11	11	11	11	11	11
11	11	11	11	11	11	11	11	11
8	8	8	8	8	8	8	8	8
14	14	14	14	14	14	14	14	14
15	15	15	15	15	15	15	15	15
8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9
11	11	11	11	11	11	11	11	11
12	12	12	12	12	12	12	12	12
12	12	12	12	12	12	12	12	12
27	27	27	27	27	27	27	27	27
13	13	13	13	13	13	13	13	13
11	11	11	11	11	11	11	11	11
20	20	20	20	20	20	20	20	20
21	21	21	21	21	21	21	21	21
11	11	11	11	11	11	11	11	11
17	16	16	17	16	16	16	16	
19	19	19	19	19	19	19	19	19
13	13	13	13	13	13	12	13	
12	12	12	12	12	11	12	11	11
22	22	22	22	22	22	22	22	22
11	11	12	11	11	11	11	11	11*
19	19	19	19	19	19	19	19	19
21	21	21	21	21	21	21	21	21
14	14	14	14	12	12	12	12	12
15	15	15	15	14	14	14	14	14
15	16	16	15	15	15	15	15	15
16	16	16	16	16	16	16	16	16
		23		23	23			
		25		25				
12	12	12	12	12	12	12		
13	13		13	13	13	13		
31	31		31	31	31	31		
21	21		21	21	21	21		
8	8	8	8	8	8	8		
25	25		25	25	25	25		
9	9	9	9	9	9	9		
21	21	21	21	21	21	21		
21	21	21	21	21	21	21		
11	11	11	11	11	11	11		
19	19	19	19	18	18	19		19
17	18	18	18	18	18	18		18
8	8	8	8	8	8	8		
13	13	13	12	13	13	13		13
35	35		35	35				
29	29		29	29	29	29		
16	16		16	16	16	16		
14	14		14	14	14	14		
33	32	34	33	33	33	33		
31	32	32	32	32	32	32		32**
36	36	36	35	36	35	36		36**

Common mutations at **DYS449**, and also in the ringed box below at **DYS570**

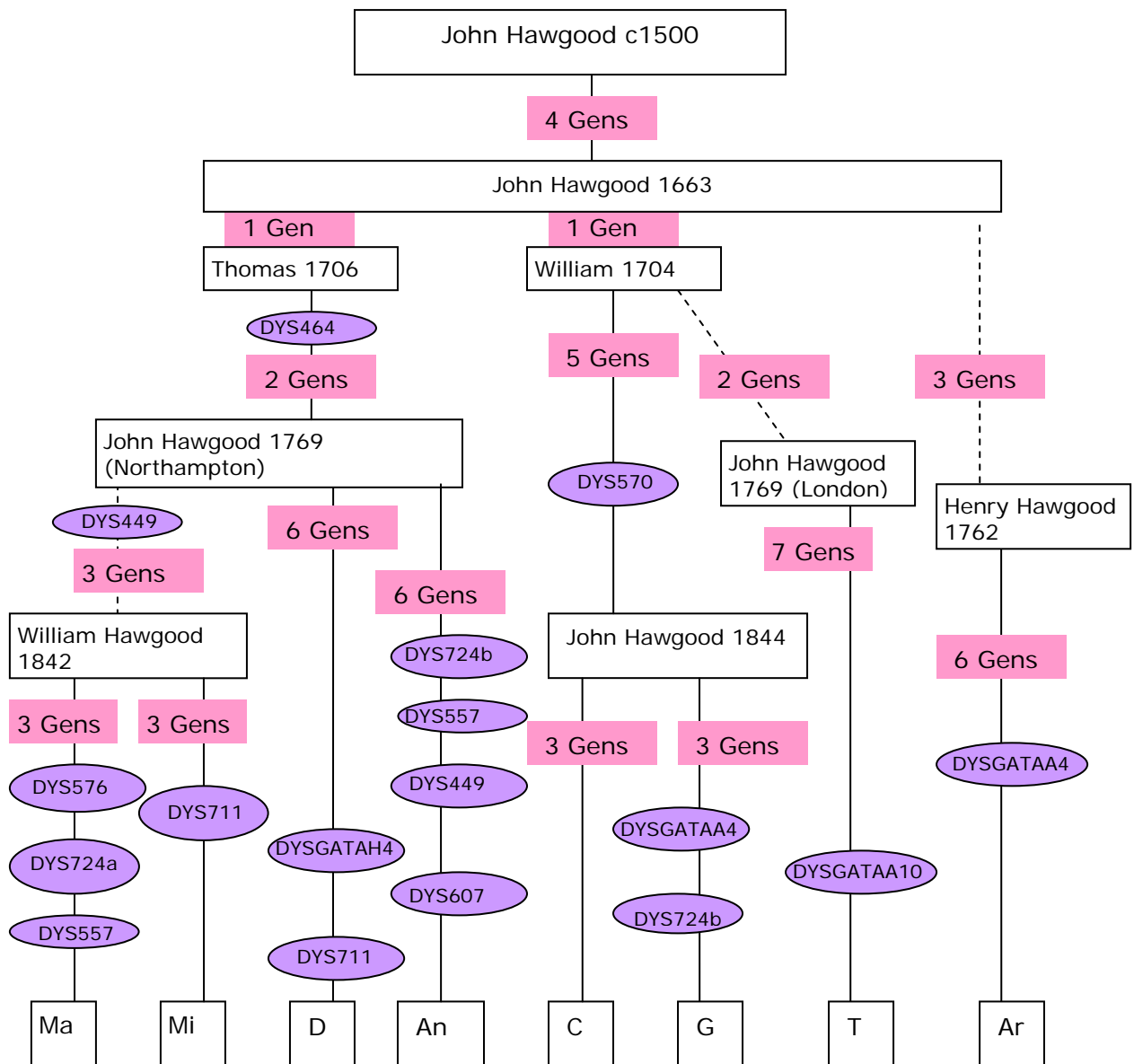
This is the most significant area where four share virtually the same values, which differ from the other four. This shows two distinct branches

These are all extraordinarily fast mutating markers, and therefore not surprising to see so many mutations

Diagram of links

The diagram below summarises the for those tested in the Hawgood family DNA study. The test results have enabled us to connect broken branches into the main family tree.

- The dotted lines show common ancestors derived from the results
- The unbroken lines show known ancestors
- Pink text boxes show generation gaps (Gens)
- Purple ovals show which markers mutated and in which period



Generation gaps - actual versus predicted

This next analysis compares the results expected generation gap with the actual generation gap, and where the actual is not known, predicts the gap using expected mathematics. Where only the 67 marker test has been taken, we have used the standard 44 marker suite to replicate the results for this test using the 67 marker results.

There are three columns of data

- 1) The actual number of mutations between individuals
- 2) The expected generation gap, which is based on the actual number of mutations and is a mathematical calculation. A forecasting table is shown in Appendix B
- 3) The known generation gap.

The values in pink show where the predicted gap matches the known gap.

So as an example, the between M and M2, there is 1 mutation in 44 markers. The predicted generation gap is 3-6 generation, and the known gap is 3 generations. The theory matches the known difference. With the 67 marker test however there are 3 additional mutations which predict a gap of 7-8 generations, which is longer than the known difference of 3 generations.

This increased variance is more likely to occur where a number of faster mutating markers are included in tests, such as the additional 23 markers in the 67 marker test compared to the 44 marker test, and it therefore becomes more likely to see sometimes unusual results.

Whilst the odd aberration will exist, it is interesting to see overall the accuracy of the mathematical model behind the predictions does correctly predict the generation gap the majority of the time. Compare this however to many online modelling tools, the predicted generation gap rarely matches as they do not use up to date values of mutation rates.

This exercise also shows the inconsistencies in individual tests as Ma and Mi are the same distance from D, yet the prediction of the generation gap between Ma & D is forecast at 12-14 generations whereas between Mi and D is 6-7 generations. Mi is the same distance from 'An' and from D, yet the forecast gap between Mi and D is 6-7 generations (the known gap is 6), and the forecast gap between Mi and 'An' is 11 generations. This is purely due to extra random mutations, but does throw out the calculation. 'An' and G have the same mutation on *DYS724b* although this is coincidence as we know that they have occurred independent of each other, and this result masks two extra mutations.

44 Test - Number of mutations

	Ma	Mi	D	An	C	T	G	A
Ma	-	1	3	2	2	3	3	2
Mi	1	-	2	3	2	2	2	1
D	3	2	-	3	1	2	2	2
An	2	3	3	-	2	3	3	2
C	2	2	1	2	-	1	1	1
T	3	2	2	3	1	-	2	2
G	3	2	2	3	1	2	-	0
A	2	1	2	4	1	2	0	-

44 Test - Expected generation gap

	Ma	Mi	D	An	C	T	G	A
Ma	-	3-6	11-14	7-10	7-10	11-14	11-14	7-11
Mi	3-6	-	7-10	11-14	7-10	7-10	7-10	3-6
D	11-14	7-10	-	11-14	3-6	7-10	7-10	7-11
An	7-10	11-14	11-14	-	7-10	11-14	11-14	7-10
C	7-10	7-10	3-6	7-10	-	3-6	3-6	3-6
T	11-14	7-10	7-10	11-14	3-6	-	7-10	7-11
G	11-14	7-10	7-10	11-14	3-6	7-10	-	1-2
A	7-11	3-6	7-11	7-10	3-6	7-11	1-2	-

Known generation gap

	Ma	Mi	D	An	C	T	G	A
Ma	-	3	6	6	8	?	8	?
Mi	3	-	6	6	8	?	8	?
D	6	6	-	6	9	?	9	?
An	6	6	6	-	9	?	9	?
C	9	9	9	9	-	?	3	?
T	?	?	?	?	?	-	?	?
G	9	9	9	9	3	?	-	?
A	?	?	?	?	?	?	?	-

67 Test - Number of mutations

	Ma	Mi	D	An	C	T	G
Ma	-	4	6	6	5	5	7
Mi	4	-	3	6	3	3	5
D	6	3	-	6	3	3	5
An	6	6	6	-	5	5	5
C	5	3	3	5	-	2	2
T	5	3	3	5	2	-	4
G	7	5	5	5	2	4	-

67 Test - Expected generation gap

	Ma	Mi	D	An	C	T	G
Ma	-	7-8	12-14	11	9-10	9-10	13-14
Mi	7-8	-	6-8	11	5-6	5-6	9-10
D	12-14	6-7	-	11	6-8	6-7	11-12
An	11	11	11	-	9-10	9-10	9-10
C	9-10	5-6	6-8	9-10	-	3-4	3-4
T	9-10	5-6	6-8	9-10	3-4	-	7-8
G	13-14	9-10	11-12	9-10	3-4	7-8	-

Known generation gap

	Ma	Mi	D	An	C	T	G
Ma	-	3	?	6	8	?	9
Mi	3	-	6	6	8	?	9
D	6	6	-	6	8	?	9
An	6	6	6	-	9	?	9
C	9	9	9	9	-	?	3
T	?	?	?	?	?	-	?
G	9	9	9	9	3	?	-

D is tested against common 56 markers

Denotes where the expected generation gap matches the known generation gap

Appendix A - Forecasting table

The table below can be used to estimate the generation gap based on the number of mutations, shown in pink highlight. The rates shown are in percentages, so 0.27% is the same as 0.0027.

Statistically likely generation gaps based on actual mutations

Rate	0.27	0.41
Markers	43	63

Generations	Expected mutations	Expected mutations
1	0.2	0.5
2	0.5	1.0
3	0.7	1.5
4	0.9	2.1
5	1.2	2.6
6	1.4	3.1
7	1.6	3.6
8	1.9	4.1
9	2.1	4.6
10	2.3	5.2
11	2.6	5.7
12	2.8	6.2
13	3.0	6.7
14	3.3	7.2
15	3.5	7.7
16	3.7	8.3
17	3.9	8.8

Actual mutations

	1	2	3	4
1	0	0	0	0
2	0	0	0	0
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	2	2	2	2
8	2	2	2	2
9	2	2	2	2
10	2	2	2	2
11	3	3	3	3
12	3	3	3	3
13	3	3	3	3
14	3	3	3	3
15	3	3	3	3
16	4	4	4	4
17	4	4	4	4

	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2
4	2	2	2	2	2	2	2	2
5	3	3	3	3	3	3	3	3
6	3	3	3	3	3	3	3	3
7	4	4	4	4	4	4	4	4
8	4	4	4	4	4	4	4	4
9	5	5	5	5	5	5	5	5
10	5	5	5	5	5	5	5	5
11	6	6	6	6	6	6	6	6
12	6	6	6	6	6	6	6	6
13	7	7	7	7	7	7	7	7
14	7	7	7	7	7	7	7	7
15	8	8	8	8	8	8	8	8
16	8	8	8	8	8	8	8	8
17	9	9	9	9	9	9	9	9

Generation gap

For example, in a 43 Markers test with 3 mutations, at a 0.27% mutation rate, the generation gap is likely to be between 11 and 15 generations.

Appendix B - Individual marker mutation rates

Summary average data

Individual data surveys

Genebase Marker test	Average	Median	Individual data surveys													
			1	2	3	4	5	6	7	8	9	10	11	12	13	
			Janzen (Jun 2008)	Chandler	www.cstl.nist.gov (2007)	McDonald	SMGF.org 7976 dataset	YHRD.org	Sorenson	Burgarella 2010	ASHG 2004	Bioinformatics	Ballantyne etc 17 markers 2009	Ballantyne etc 186 markers Sep 2010	Worldfamilies (McD rates)	
DYS19	44	0.0023	0.0023													
DYS385a	44	0.0032	0.0028	0.0023	0.0015	0.0025	0.0016	0.0015	0.0023	0.0015		0.0030	0.0027	0.0044	0.0016	
DYS385b	44	0.0034	0.0033	0.0023		0.0021	0.0033	0.0056	0.0021	0.0057		0.0034	0.0021	0.0021	0.0033	
DYS388	44	0.0005	0.0005		0.0006		0.0005	0.0004		0.0002			0.0008		0.0005	
DYS389i	44	0.0025	0.0022		0.0019	0.0024	0.0021	0.0022	0.0025	0.0022			0.0020	0.0028	0.0021	
DYS389il	44	0.0028	0.0028		0.0024	0.0035	0.0028	0.0026	0.0036	0.0027		0.0011	0.0032	0.0031	0.0028	
DYS390	44	0.0034	0.0032		0.0031	0.0025	0.0045	0.0044	0.0021	0.0053		0.0034	0.0028	0.0022	0.0045	
DYS391	44	0.0027	0.0029		0.0027	0.0028	0.0036	0.0032	0.0026	0.0041		0.0011	0.0010	0.0030	0.0036	
DYS392	44	0.0014	0.0012		0.0005	0.0008	0.0016	0.0015	0.0041	0.0016		0.0005	0.0023	0.0008	0.0016	
DYS393	44	0.0015	0.0014		0.0008		0.0014	0.0011	0.0010	0.0014		0.0026	0.0023	0.0018	0.0012	
DYS426	44	0.0003	0.0003		0.0001		0.0005	0.0003				0.0005	0.0001	0.0000	0.0005	
DYS437	44	0.0017	0.0017		0.0010		0.0020	0.0018	0.0012	0.0021		0.0023	0.0014	0.0013	0.0020	
DYS438	44	0.0008	0.0008		0.0006	0.0007	0.0012	0.0010	0.0003			0.0007	0.0012	0.0006	0.0012	
DYS442	44	0.0050	0.0041		0.0032							0.0019	0.0049	0.0098		
DYS445	44	0.0026	0.0025	0.0030				0.0030				0.0025	0.0024	0.0022		
DYS446	44	0.0035	0.0031	0.0037			0.0032	0.0031				0.0007	0.0075	0.0027		
DYS447	44	0.0030	0.0029		0.0026		0.0045	0.0031		0.0040		0.0007	0.0023	0.0021	0.0045	
DYS448	44	0.0017	0.0016		0.0014	0.0011	0.0028	0.0024	0.0016			0.0016	0.0011	0.0011	0.0028	
DYS449	44	0.0088	0.0081		0.0084		0.0056	0.0065		0.0078		0.0096	0.0124	0.0122	0.0075	
DYS453	44	0.0022	0.0022									0.0022				
DYS454	44	0.0007	0.0005	0.0002			0.0005	0.0002		0.0005		0.0021			0.0005	
DYS455	44	0.0007	0.0005	0.0002			0.0005	0.0003		0.0005		0.0021			0.0005	
DYS456	44	0.0054	0.0051		0.0074	0.0053	0.0053	0.0042		0.0033	0.0083		0.0044	0.0049		
DYS458	44	0.0073	0.0066	0.0081		0.0106	0.0066	0.0058	0.0064	0.0063	0.0048	0.0090		0.0084	0.0066	
DYS459a	44	0.0021	0.0024		0.0013		0.0014	0.0024		0.0026				0.0027		
DYS459b	44	0.0021	0.0024		0.0013		0.0014	0.0024		0.0026				0.0027		
DYS460	44	0.0034	0.0028		0.0040		0.0028	0.0029			0.0025	0.0023		0.0062	0.0028	
DYS461	44	0.0023	0.0026				0.0028	0.0023			0.0030			0.0010		
DYS462	44	0.0015	0.0011				0.0005	0.0005			0.0028	0.0011		0.0027		
DYS468	44	0.0017	0.0017											0.0017		
DYS484	44	0.0027	0.0027								0.0028			0.0026		
DYS522	44	0.0026	0.0023				0.0045				0.0023			0.0010		
DYS527a	44	0.0065	0.0065											0.0065		
DYS527b	44															
DYS531	44	0.0012	0.0010	0.0004							0.0023			0.0010		
DYS557	44	0.0034	0.0034	0.0032			0.0036				0.0031			0.0038		
DYS588	44	0.0004	0.0004								0.0004			0.0004		
GATAA10	44	0.0034	0.0038	0.0038			0.0045	0.0038			0.0030	0.0011		0.0033	0.0045	
GATAA4	44	0.0048	0.0048	0.0048	0.0048	0.0061	0.0045	0.0042	0.0052	0.0049	0.0029	0.0068	0.0056	0.0038	0.0045	
GATAC4	44	0.0034	0.0035	0.0024		0.0046	0.0028	0.0023	0.0035		0.0028	0.0045	0.0039	0.0039		
GATAH4	44	0.0032	0.0031		0.0029	0.0043	0.0036	0.0030	0.0024		0.0022	0.0034	0.0031	0.0032	0.0036	
YCAIIa	44	0.0016	0.0014		0.0012		0.0014	0.0025							0.0014	
YCAIIb	44	0.0016	0.0014		0.0012		0.0014	0.0025							0.0014	
DYS464a	44	0.0050	0.0046		0.0057		0.0035							0.0073	0.0035	
Average - 44	0.0028	0.0027	0.0028	0.0025	0.0034	0.0027	0.0027	0.0028	0.0033	0.0024	0.0040			0.0036	0.0028	

DYS413a	67	0.0020	0.0020	0.00202												
DYS413b	67															
DYS436	67	0.0003	0.0004	0.00018							0.00044			0.0004		
DYS444	67	0.0029	0.0026	0.00321			0.00224	0.003			0.002	0.0012		0.00545		
DYS452	67	0.0033	0.0040					0.00174			0.004			0.00402		
DYS463	67	0.0017	0.0016				0.00204	0.0016			0.0007	0.0025		0.00151		
DYS464b	67	0.0050	0.0046		0.00566		0.0035							0.00727	0.0035	
DYS464c	67	0.0050	0.0046		0.00566		0.0035							0.00727	0.0035	
DYS464d	67	0.0055	0.0057		0.00566		0.0035							0.00727		
DYS472	67	0.0002	0.0002	0.00001							0.0004					
DYS481	67	0.0058	0.0054	0.00544							0.0069			0.00497		
DYS511	67	0.0017	0.0015	0.00128							0.0024			0.00152		
DYS518	67	0.0184	0.0184											0.0184		
DYS520	67	0.0024	0.0024	0.00245			0.00216				0.0023			0.0027		
DYS537	67	0.0019	0.0024	0.00057							0.0028			0.0024		
DYS570	67	0.0075	0.0079	0.00838	0.0079		0.004778				0.0042			0.0124		
DYS576	67	0.0086	0.0079		0.01022		0.00552				0.0042			0.0143		
DYS590	67	0.0005	0.0005	0.00054							0.0004					
DYS607	67	0.0039	0.0039		0.00411						0.0037					
DYS612	67	0.0145	0.0145											0.0145		
DYS614	67	0.0043	0.0043											0.00432		
DYS644	67	0.0032	0.0032											0.00322		
DYS710	67	0.0175	0.0175	0.0175												
DYS711	67															
DYS724a	67	0.0353	0.0353		0.03531											
DYS724b	67															
Average - 67	0.0041	0.0040														

- For DYS724 and 710/11 mutation rates may be over both markers in each set - hence for prudence, only once value is used

- DYS 464 ignored for average rate calculation

- DYS413a FTDNA at <https://docs.google.com/viewer?a=v&pid=explorer&chrome=true&srcid=0By9Y3jb2fORNY2ZJZWM4OGHtZjY2Y0NDQwLWlyYzZmZmUwY2ZlYjFiZi>

Mutation rate sources

No	Sample Size	Source
1	8430	Chandler values taken from http://www.timjanzen.com/variance_calculator.xls where not in jogg.info/22/chandler
2	15295	http://www.jogg.info/22/Chandler.pdf http://www.jogg.info/22/Chandler.htm
3	5000	http://www.cstl.nist.gov/biotech/strbase/ystr_fact.htm
4	500	McDonald (2004-6) Sample size estimate
5	7,976	http://www.smfg.org/resources/papers/ASHG2004-4.pdf Sample size estimate
6	6364-25306	http://www.yhrd.org/Research/Loci
7	5,000	http://www.smfg.org/resources/papers/ASHG2004-4.pdf (Ysearch.org dataset) Sample size estimate
8	80-13948	https://sites.google.com/site/navascuesresearch/publications-conferences/journalpublications/burgarellanavascues2010
9	864	http://www.smfg.org/resources/papers/ASHG2004-3.pdf
10	3780	http://bioinformatics.oxfordjournals.org/content/26/18/1440.full.pdf+html
11	3384-11900	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766043/table/Tab2/
12	1700	http://www.cell.com/AJHG/supplemental/S0002-9297(10)00419-2 and at http://www.sciencedirect.com/science/article/B8JDD-50XJR
13		http://www.worldfamilies.net/marker
Total circa	74045	