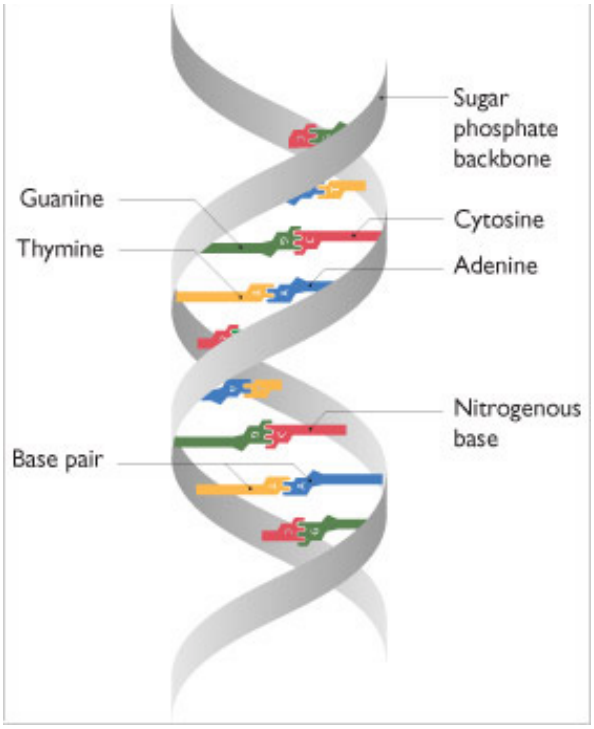
DNA and its Uses in Genealogy

This paper aims to provide a simple explanation of DNA and how it can be used in genealogy. It is meant as an introduction. It uses examples from the **Warburton DNA Project**. and **Warburton One-name Study** to aid understanding. There are references that will provide a more professional and detailed explanation.

What is DNA?

The following description and diagram of DNA are taken from the Genetics Home Reference website at *http://ghr.nlm.nih.gov/handbook/basics/dna*:

> DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. Nearly every cell in a person's body has the same DNA. Most DNA is located in the cell nucleus (where it is called nuclear DNA), but a small amount of DNA can also be found in the mitochondria (where it is called mitochondrial DNA or mtDNA).
>
> The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people. The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences.



> DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is also attached to a sugar molecule and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix. The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.
>
> An important property of DNA is that it can replicate, or make copies of itself. Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of

bases. This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell.

Strands of DNA are organised into genes. A gene could be defined as the shortest string of DNA that actually does something useful in our development. However between the genes are strings of useless or junk DNA that do nothing (maybe they did once in earlier stages of evolution). These useless strings of bases are important for genealogy but I will come back to that later.

Genes and their attendant junk DNA are organised into chromosomes (so called because scientists use coloured dyes to identify them). The human genome (i.e. our DNA) consists of 46 chromosomes, or rather 23 pairs. We inherit one set of 23 from our father and one from our mother. When we come to pass on 23 chromosomes to our children each chromosome is a composite containing some DNA from each partner in the original chromosome pair. Thus the DNA we pass to our children is a mixture, some from our father and some from our mother. The mixture is different every time.

However one of the pairs of chromosomes is different. This is the pair of so-called sex chromosomes, the X and Y-chromosomes which determine the sex of a child. A female has 2 X-chromosomes, one from each parent. However a male has an X-chromosome from his mother and a Y-chromosome from his father. It is a gene on the Y-chromosome that causes a baby to be a boy. In the absence of this gene the default is always to produce a girl.

The significance of this is that (unlike all the other chromosomes) the Y-chromosome is never mixed with a copy from the mother. It passes unchanged from father to son through the generations.

Now the body is very good at faithfully copying DNA from generation to generation, but it is not perfect (otherwise evolution wouldn't work). Very occasionally a copying mistake occurs. For example an adenine (A) base may become a thymine (T) base. If it happens in a gene it may cause disease, or rarely it improves the gene. But if it happens in junk DNA it has no effect and so the mistake continues to be copied from generation to generation. It is these differences that make DNA useful in historical and genealogical studies.

There is one other piece of DNA that is passed unchanged from generation to generation. It is in addition to the 46 chromosomes and acts as the energy source for a cell. It is called mitochondria and is only passed down the female line. Males do have it, inherited from their mother, but don't pass it on.

DNA Testing for Genealogy

There are currently four types of DNA test marketed for people interested in their genealogy or genetic history. These are two forms of Single Nucleotide Polymorphism (SNIP) tests, Short Tandem Repeat (STR) tests and autosomal tests. I will describe each of these in the following sections, illustrating each with my own test results. One further note about terminology, the locations of interest in comparing test results are referred to as markers, whether we are examining single bases as in SNIP tests, or longer sequences as in STR or autosomal tests.

The DNA test itself is the same in all cases and is actually very simple. The test kit is mailed to you. It consists of a couple of cotton buds and a return envelope. You wipe the inside of your mouth with the cotton buds and mail them back. This sample can then be used for any or all of the tests described. Results are returned within a few weeks. The sample may be retained by the testing company so that additional tests can be ordered without the need for a new test kit.

Single Nucleotide Polymorphism

I mentioned above that very occasionally a DNA letter is copied wrongly e.g. an A becomes a T. This is known as a Single Nucleotide Polymorphism or SNP or SNIP. So a SNIP is when a letter is copied wrongly.

SNIPs occur very rarely. A specific letter may have only changed once in the whole of modern mankind's existence (150-180,000 years). The global population can therefore be divided into those who have the SNIP and those who don’t. Of those who do, a proportion may also have a later SNIP. By using a small number of SNIPs and understanding their sequence, their current geographical spread, and the amount of change at various locations, researchers have been able divide the world population into a small number of clans, called haplogroups in the jargon,

determine where and when these haplogroups originated, and piece together ancient human migration patterns.

Over time as more SNIPs are discovered these migration patterns can be given more detailed and more recent definition. The haplogroups are divided into clades and sub-clades using genetic family trees called phylogenetic trees. Of course this has to be linked to the archaeological evidence to be meaningful.

We have seen there are two pieces of DNA which are passed unchanged from generation to generation, the Y-chromosome and mitochondria. Thus we have SNIP tests for the Y-chromosome and for mitochondria. The main difference between these tests is that we all carry mitochondria and so can take the mitochondrial SNIP test, but only males have a Y-chromosome and can take a Y-chromosome SNIP test.

By performing exactly the same haplogroup analysis on both the Y-chromosome and mitochondria two separate, corroborating pictures of mankind's past can be built. The results can be fascinating. For example my mitochondria, and my Y-chromosome have completely different histories which are discussed below.

Mitochondrial SNIP Tests

My very first DNA test was a mitochondrial test. I was introduced to the subject by a book called **Out of Eden** by Stephen Oppenheimer. The subject of the book was what I now know to be a new 'science' called phylogeography. This combines phylogenetics with traditional archaeology to study the ancient migrations of peoples. It focused primarily on mitochondria, though the Y-chromosome was included. The startling conclusion of the book was that all non-Africans in the world are descended from a small group of humans that left Africa 80,000 years ago.

I then came across two books by Professor Bryan Sykes, **The Seven Daughters of Eve**, and **Adam's Curse**. These books have very readable discussions of the science, and lots of interesting anecdotes.

In summary mitochondria has been used to divide the world population into 36 haplotypes, 13 in Africa. Ninety-five percent of Europeans fall into 7 haplotypes which are discussed in detail in **The Seven Daughters of Eve**.

Perhaps the most startling assertion, though it is logical when you think about it, is that each mitochondrial clan must be descended from a single woman. The copying error occurred just once, so everyone carrying the error must be descended from the first woman to carry the error. Not only that, but by linking the clans in a tree, one clan becomes the source of all the others. Not unnaturally the origin of this clan is called Eve. This doesn't mean that Eve was not one of a population of similar early humans. It's just that no descendants of her contemporaries exist today.

Professor Sykes has set up a testing company called **Oxford Ancestors** to allow people to get their DNA tested, so back in 2006 I got my first test test.

Mitochondria is a small piece of DNA having approximately 16,000 bases. Of these 2 regions, called Hyper Variable Regions, or HVR1 and HVR2 are used to define haplogroups.

HVR1, containing about 400 bases, is sufficient to determine your haplogroup and this is what **Oxford Ancestors** tested. I had 3 SNIPs reported in this sequence. The one at position 126 is found in two haplogroups, J and T, but that at position 69 is specific to haplogroup J. The one other SNIP, at position 366, put me in the main group of Js, not in one of a number of sub-clans that had been defined.

J is a complex haplotype with several defined sub-groups. The other six main haplogroups found in Europe are all believed to have existed, or derived from groups that existed in Europe before the last Ice Age. Mitochondrial Clan J however originated in the Near East (possibly the Caucasus) and only moved into Europe after the last Ice Age when Neolithic farmers began to move into Europe 10,000 years ago. It followed 2 tracks, one of which followed the coast around the Mediterranean, and eventually up to Britain, whilst the other moved through central Europe.

The presence of haplogroup J in Europe, but in relatively small numbers (10% of the population) alongside the descendants of the earlier hunter-gatherer population answers an old historical argument. Did the gradual adoption of farming across Europe represent the migration of an idea, or the migration of farming peoples who replaced the indigenous population? The answer lies between the two. There was a migration of people, but they did not replace the existing population

and the idea of farming spread into the indigenous population. My SNIP at position 366 was not found in the Near East so appears to have originated in Europe.

More recently (2013) I took the **Geno 2.0** test. This is used by **The Genographic Project** run by **National Geographic**. My main objective was the Y-chromosome SNIP test which I will discuss later, but it includes both a mitochondrial SNIP test and an autosomal test (all for $200).

The **Geno 2.0** test covered both HRV1, and HRV2 an additional 576 bases. My result classified me as J1c2 and suggested that J1c originated in Europe about 13,000 years ago.

I have not yet studied this result in detail or identified the SNIPs that define my classification. However one feature of the **Geno 2.0** test is you can transfer your result, and your sample to **Family Tree DNA** the company that hosts my **Warburton DNA Project**. In fact I believe the same laboratory does testing for **Family Tree DNA** and **The Genographic Project**. **Family Tree DNA** host a large number of special interest projects, including surname projects and haplogroup projects. I have recently joined the **J-mtDNA** project and hope to understand my result better in due course.

Y-chromosome SNIP Tests

As with mitochondria I first took a Y-chromosome SNIP test in 2006, this time with **DNA Heritage**, and then repeated it recently (2013) as part of the **Geno 2.0** test. The Y-chromosome is much larger than mitochondria and so offers many more SNIPs. In fact new SNIPs are constantly being discovered.

My **DNA Heritage** result classified me as R1b3*. However the R haplogroup has been reclassified more than once and this is now classified as R1b1a2. I am defined as R1b1a2 because I have SNIPS at markers M207 (which defines me as haplogroup R), M343 (which makes me R1b), and M269 (which defines me as R1b1a2).

Haplogroup R1b is the dominant Y-chromosome in Western Europe, particularly along the Atlantic seaboard. Most of these are in sub-clade R1b1a2 which includes 40-70% of the population of continental Western Europe rising to 81% in the Basque country, 85% in Ireland, and over 90% in parts of Wales.

This distribution led researchers to conclude that members of the R1b haplogroup are descendants of the first modern human migrants into Europe some 35-40,000 years ago. This is known as the Upper Palaeolithic migration and was characterised by the Aurignacian culture. During the last Ice Age they retreated to a number of refuges in southern Europe. The large proportion of R1b in the Basque region led to the belief that the mutation that defines R1b occurred in the Iberian refuge.

Since 2010 new research has concluded that R1b arose in Central Asia because the oldest sub-clades are found there, whilst those in Western Europe are much newer. The exact time it arrived in Europe is still uncertain but it was probably as a result of the westward migration of Indo-Europeans. The dominant position of Rb1 in Western Europe is not believed to be the result of wholesale population replacement, but simply that men in invading elites have higher reproductive opportunities which means their Y-chromosomes soon come to dominate.

My **Geno 2.0** test added further SNIPs, the most important being at U106 and Z306 . In fact I was defined as R-Z306, but this is defined in the literature as R1b1a2a1a1c1a.

The SNIP at marker U106 is important as there is an Rb1-U106 project at **Family Tree DNA**. About 25% of R1b in Europe have the U106 SNIP, and it is most common in the Netherlands and Northern Germany. It probably originated in Austria in the Hallstatt period (5-8,000 BC).

U106 is further subdivided by additional SNIPs. Only Z306 was reported in my **Geno 2.0** results, but this means I also have Z381, Z156, and probably Z307, and Z305. Z305 seems to have occurred about 500-2500BC, probably in Northern Germany.

Y-chromosome Short Tandem Repeat Tests

So far we have discussed SNIPs. Unfortunately these occur too infrequently to help with family history. Every related male Warburton would probably have an identical profile. Fortunately there is another test which is more helpful, but it can only be carried out on the Y-chromosome.

It so happens that there are some short DNA sequences that are repeated several times. Whereas with SNIPs we were dealing with a change to a single base or letter in the DNA sequence, you can think of these sequences as words that are repeated several times. Every now and then the number of copies of the word changes. For example one may be added so, whereas there were 10 repeats before, there are now 11.

These strings of words are called Short Tandem Repeats (STRs), so a test for them is an STR test. There are a number of locations or markers where they occur on the Y-chromosome. There are tests available for different numbers of markers, and different testing companies test different markers.

I began the Warburton DNA project using a 43 marker test from **DNA Heritage**. However **DNA Heritage** ceased trading in 2011 and I switched to the **Family Tree DNA** 37 marker test, of which only 32 makers are common with the original 43 marker test. **Family Tree DNA** offer tests for 12, 37, 67, and 111 markers.

To understand which tests are most appropriate it is necessary to understand how STR tests are used. This includes concepts such as mutations, mutation rates, genetic distance, and probabilities.

The objective of STR tests is to compare results to determine if the persons tested are sufficiently matched to have a recent common ancestor. When two STR results are compared the number of markers with different values are counted. This number is known as the genetic difference . For example, the only difference between my profile and that of my genetic cousin Clive is that at marker DYS458 I have 16 repeats and he has 17, so our genetic distance is 1.

The genetic difference is used in conjunction with mutation rates to calculate a range of probabilities for the number of generations since a common ancestor lived. These are called Time to Most Recent Common Ancestor (TMRCA) calculations.

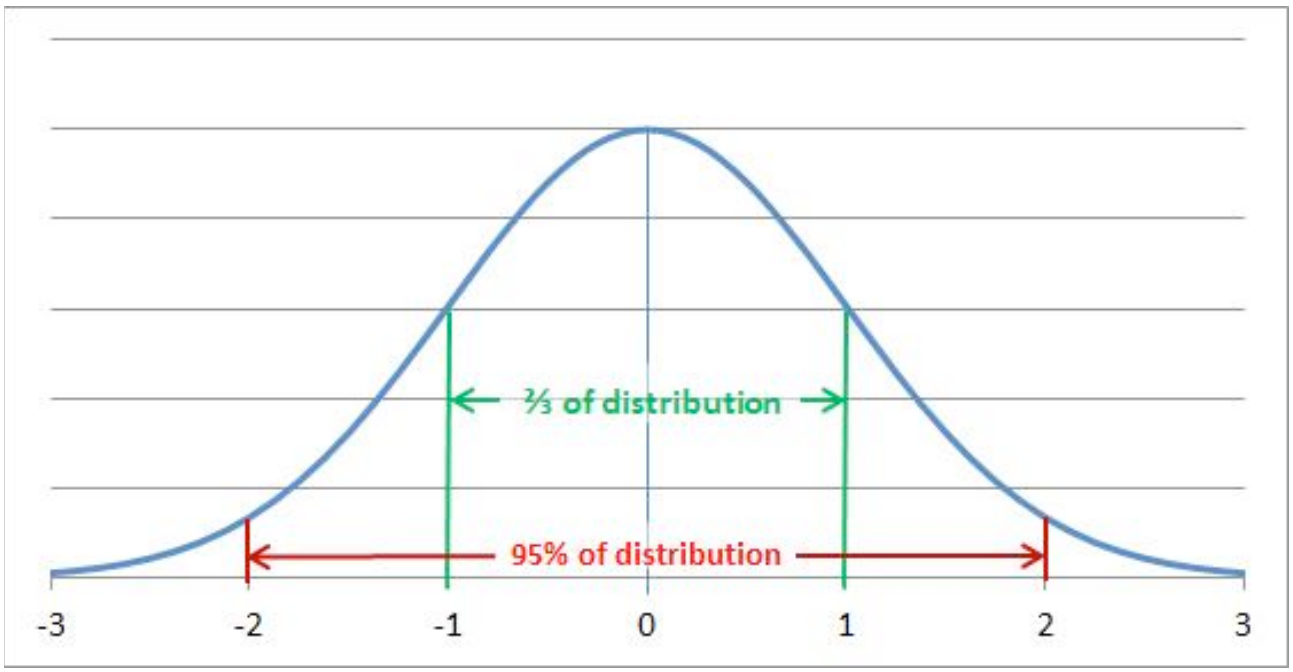
A mutation is a change in the number of repeats at a particular marker. Research is continuing on just how often these mutations occur. When I started my project the often quoted figure was one mutation in every 500 transmissions of a marker from father to son, or a rate of 0.2%. More recently I have found a number of sources that don't entirely agree but follow a pattern. Mutation rates would seem to vary considerably between markers, and maybe for a single marker between families. The 5 unique FTDNA markers are amongst the most volatile. It is possible further research will cause our knowledge of mutation rates to change again in the future.

To put mutation rates into perspective, a marker with a mutation rate of 0.3% will change its number of repeats once in every 333 transmissions of the marker from father to son. If there are 33 markers with the same mutation rate, a change should occur in one of those markers, on average once in every 10 transmissions of the Y-chromosome from father to son. This is one mutation in 10 generations, or if a man had 10 sons, one son would have a mutation at one location.

For many markers there is also documentation on the frequency at which each possible result occurs. Each possible value is known as an allele. Each allele will occur at a different frequency. A match with an allele that has a low frequency may be particularly significant. See my **Mutations Table** for details and sources on mutation rates, allele distributions. TMRCA calculations produce a range of probabilities for the time to the most recent common ancestor. For those of you with no background in statistics here is a (hopefully) simple explanation. Suppose I have a bag containing 500 balls, of which 50 are blue and the rest red. I now select 10 balls at random. What are the chances of me getting 1 blue ball? If I kept repeating the test I should get 1 blue ball more frequently than any other result, but on occasions I will get no blue balls, and on other occasions two, three, or rarely even more blue balls. By repeating the test many times the frequency of each result can be determined and plotted on a graph.

This will tell me what are the chances (the probability) of any given test returning just one blue ball, or any other number of blue balls. Of course there are mathematical formulae to work out the results without having to keep testing it.

So for any combination of the number of markers tested, the genetic distance, and the average mutation rate it is possible to calculate a probability curve for the number of generations to the most recent common ancestor. The curve would look like the one below with frequency on the vertical axis, and number of generations on the horizontal axis. The chart shows that two thirds of possible values lie within one standard deviation of the mean, and 95% within two standard deviations. Standard deviation is a calculation of the variability in the possible values.



If two results are from cousins who share a common ancestor five generations ago there are 10 Y-chromosome transmissions between them, five on each line, and this must be taken into account in the calculations. Time to Most Recent Common Ancestor (TMRCA) calculators have been developed to perform the calculation, and several are available on the web. The actual probabilities for generations to the most recent common ancestor, based on the appropriate mutation rates for each possible comparison (i.e.between 2 **DNA Heritage** results, 2 **Family Tree DNA** results, or the common marker set) are given in the **Mutations Table**.

To convert generations into time it is necessary to multiply the number of generations by the average length of a generation. There are references to a generation being 20-25 years but these are focussed on inheritance and the first born. In my own line the average of the last ten generations is over 37 years, with one of my ancestors being born when his father was 59 years old. There is probably considerable variation from line to line but I use an average of 35 years per generation to calculate the time to the most recent common ancestor.

The number of repeats for a marker can change up or down, and occasionally by more than one. It could be that due to these changes cancelling each other out, two people who are unrelated finish up with the same profile. The probability of a random match is low, but in a large population some random matches are inevitable. Therefore matches are only considered meaningful when there is additional evidence, such as a shared surname, to link two people. This is why most STR studies are surname studies, though there are some locational ones.

Surnames were introduced around 12-1300 AD, when feudal estates needed them for record keeping. In fact it is almost exactly 750 years since Sir Piers de Dutton built a manor house at Werberton and began to style himself de Werberton, At 30 years per generation this works out as 25 generations. Two modern day descendants of Sir Piers de Werberton would be 50 transmissions apart so we might expect a 37 marker test with a mutation rate of 0.42% to produce a genetic distance of six.

I began my own experience of STR testing in 2006 with a 10 marker test from **Oxford Ancestors.** This was sufficient to predict that my Y-chromosome haplogroup is R1, but wasn't much use for anything else. After more research I realised I needed more markers, and people to compare with, so I started the **Warburton DNA Project** and took a 43 marker test. I recently upgraded this by testing the 5 markers in the **Family Tree DNA** 37 marker test that weren't included in the 43 marker test.

In addition to finding matches in a surname project, STR results can also be used to give greater definition to Y-chromosome SNIP results. Even fairly recent SNIPs may have occurred a thousand or more years ago. One way to further divide the clade for such a SNIP is to use STR results. For example my most recent SNIP is Z306, and most people with this also have Z305. By looking at

certain STR markers with a low frequency allele, 60% of Z305 has been grouped into a clade called The King's Cluster. This cluster is being further divided using detailed 111 marker STR results.

I have a high expectation of belonging to this cluster because I have a 37 marker match with a group of Duttons, and one of these has proved to be a member. Remembering that the first Warburton was originally a Dutton I am now taking a 111 marker test. There is also a new SNIP which it is believed will divide the cluster and I am taking this also.

Autosomal Tests

The tests discussed so far concentrate on the DNA that comes from a single parent. Autosomal tests such as Family Finder from **Family Tree DNA** test DNA from the 22 pairs of chromosomes that contain a mix of DNA from all ancestors. Strings of DNA are compared either between two test results to seek strings that come from a common ancestor, or against reference populations to determine ethnic origins.

As a genealogical tool autosomal test can help to prove cousins with a common ancestor up to about five generation back. My **Geno 2.0** test included an autosomal test but I have no plans currently to seek matches with possible cousins. I did receive the results of reference population matching and was determined to be 41% Northern European, 40% Mediterranean, and 18% Southwest Asian. This most closely resembled the German reference population which has 46% North European, 36% Mediterranean, and 17% Southwest Asian. The British reference population has an even higher North European content than German with 50%, with just 33% Mediterranean, and 17% Southwest Asian. I guess I am more Mediterranean than the average Brit.

I am also reported to be 2.2% Neanderthal and 2.0% Denisovan. This is because early modern human populations met the descendants of earlier humanoid migrations from Africa and interbred. A small amount of genetic material from these earlier species is found in most non-African populations, suggesting it conferred advantages that resulted in the descendants of this interbreeding coming to dominate modern populations.

Using the Results in the Warburton DNA Project

The **Warburton DNA Project** uses the Y-chromosome 37 marker STR test from **Family Tree DNA**. This usually costs $149, though twice yearly promotions usually give a $20-30 reduction. It provides sufficient markers to be able to determine matches in a surname study. The 12 marker test is too few to be sure of a match, though it will eliminate the possibility of a match in many situations. Since the 12 marker test is now quite cheap at $49 it could be used to determine if a match is possible before deciding on a 37 marker upgrade. The 67 and 111 marker tests are more specialist tests that could provide additional information to narrow the time to a most recent common ancestor, refine a phylogenetic tree, or be used to add definition to the clades of the most recent SNIPs.

A result consists of a string of numbers, one for each marker, and a prediction of the haplotype that would result from a Y-chromosome SNP test. Each of the markers has a name, such as DYS19. The associated number is typically between 8 and 30, and is the number of times a small sequence of DNA is repeated at the location of that marker.

The first step is to look for matches with previous results. Most results are either clear matches, with a genetic distance of 5 or less from another result, or a mismatch with a genetic distance of 9 or more from all other results. I will refer to a group of matched results, or a single unmatched result, as a DNA profile.

The following table shows the probability for the number of generations to the most recent common ancestor given the genetic distance between 2 results on the **Family Tree DNA** 37 marker test.

Probability	10%	25%	50%	75%	90%

Genetic Distance	Generations to Most Recent Common Ancestor				
0	1	1	2	5	8
1	2	3	6	9	13
2	4	6	9	13	18
3	6	9	13	17	23
4	9	12	16	22	27
5	11	15	20	26	32

For example I have a genetic distance of 4 from Mark. There is an 80% chance that our most recent common ancestor lived between 9 and 27 generations ago, with mean at 16 generations (480 years). However there is still a 20% chance he lies outside these boundaries.

The first genetic cousin I discovered was Clive. We tested at **DNA Heritage** so the mutation rate is different (see the **Mutations Table** for a **DNA Heritage** version of the above table). We have a genetic distance of 1 on the 43 marker test, so our common ancestor is most likely between 3 and 20 generations back, with the mean at 9 generations. In fact, based on genealogy, our most likely common ancestor is 9 generations back. See **My Genetic Links** for more about my search for links to my genetic cousins.

Non-Paternal Events

To date the success rate in finding matches is about 50%. This proportion may increase as the number of tests increases. The failure to match arises because, whereas DNA profiles always pass from father to son ,this isn't always the case with surnames. There are occasions when a male receives the Warburton name from someone other than his biological father. The rate of such events is apparently about 2% per generation.

There are a number of reasons why a son might not take his natural father's name. Infidelity and illegitimacy are obvious reasons, but it isn't unheard of for a family to take the wife's name if she brings a considerable inheritance to the family. For example the Egerton family who inherited Arley Hall in 1813 from Sir Peter Warburton through his niece, changed their name to Egerton Warburton.

Also a name might simply be adopted, for example because it is a step-father's name, or it is politically expedient. There is a case documented in the London Gazette of 1792 of a Charles Terence Mongon adopting the name Warburton (the name of his maternal cousins) apparently because it would aid his preferment in the protestant Church of Ireland.

Of course every Warburton line begins with a 'non-paternal event', though except where we have documented history, like the baptism record of an illegitimate child, or the evidence that Sir Piers de Dutton adopted the Warburton name when he built a house at Warburton, we may never be sure of the exact details.

Triangulation

When the results of two participants in the **Warburton DNA Project** match this identifies a triangle in which the participants' common ancestor is at the apex, the participants form the base, and all the other descendants of the common ancestor are contained within it. All the male Warburtons in this triangle would have the same DNA profile as the first two participants, unless of course there was a further 'non-paternal event' in their line from the common ancestor.

In the **Warburton One-Name Study** I am using traditional genealogical research to build a series of family trees. These family trees are the inverse of a traditional family tree in that they follow multiple lines of descent from a single common ancestor, so I refer to them as clans. As such they mirror the triangles produced from DNA matches. I have traced my own line back to an earliest ancestor called George who died in 1639. My clan tree (the Hale Barns clan) traces all known lines of descent from George. Of course there may be more lines I have yet to uncover.

My aim is to assign a DNA profile to each of the Warburton clans, that I document. A single Y-chromosome DNA result will link a DNA profile to a Warburton clan, but it will give no clue as to

how long the profile and the clan have been linked. Clues may come from genealogy. Many illegitimacies are recorded and so would give a probable time of linking. However the only way to get any clarity is to find a matching profile, and determine the common ancestor, or at least when he might have lived. It is not always possible to be precise, but boundaries can be determined.

Whenever an unmatched profile is obtained my objective is to achieve a match to enable triangulation. One strategy is to hope a match turns up. However, given the current rate of testing this might take a while. The alternative is to look where a match should be. If the unmatched profile is linked to a family tree of sufficient breadth it may be possible to identify a distant cousin who should be a match, and to test him. Focussed testing like this might produce quicker results, though if a 'non-paternal event' intervenes it might take a third or fourth test to isolate the problem. When a match is found it might not be to the original participant, who would now know his link to the clan is through a 'non-paternal event'.

Where a match is found and a triangle created this triangle can be fitted onto the Warburton clan family trees. A triangle may cover part of a clan where two participants are from that clan and their common ancestor is a descendant of the clan's earliest ancestor. We would not be certain at this stage if we had found DNA profile of the whole clan, or part of a segment of the clan which is defined by a 'non-paternal event'.

When the match is with a member of a different clan then the two clans are linked by a common DNA profile. The common ancestor may be within the clan with the deepest known history, making the younger clan a new branch of the older one. We believe this to be the case with my genetic cousin Clive. His Ringley clan is now considered to be a branch of the Hale Barns clan.

However it might be that the two clans are descended from an earlier common ancestor who pre-dates the earliest known ancestor in each clan. This would be the case if the earliest ancestors of each clan were contemporary, and clearly not related.

Where a match indicates a high probability that a most recent common ancestor lived since 1600 AD then it is worth using genealogical tools such as parish records to find a link. **My Genetic Links** describes my search for for such links to my genetic cousins.

Even when the actual links cannot be determined there may be geographic links. For example the Hale Barns clan and its associated clans seem to have origins in North Cheshire, close to the village of Warburton itself. There are two other small groupings, one located in Lancashire in the Haslingden-Bury area, and one distinct clan originating in Ireland, though possibly linked to the descendants of Sir Piers de Warburton, if we accept family legend. It may be particularly appealing to overseas Warburtons to see if they can link to one of these groups and so establish their geographic origins more closely.

Once a triangle is defined there is little point in anyone else from within the triangle being tested as he should match the existing results. There might be a small benefit from refining the phylogenetic tree (see below), but this could be offset by the risk of uncovering an unknown, recent 'non-paternal event'.

One of my objectives for the project is to determine whether modern Warburtons are descended from one original adopter of the name (i.e. Sir Piers de Werberton), or from a number of separate adopters. Given an assumed 'non-paternal event' rate of 2% per generation, and 25 generations (750 years) since Sir Piers then about 50% of modern Warburton will have inherited their Y-chromosome from Sir Piers, and 50% will not have. So far, the largest group of matched results is just below 30% of the total so the question has yet to be resolved.

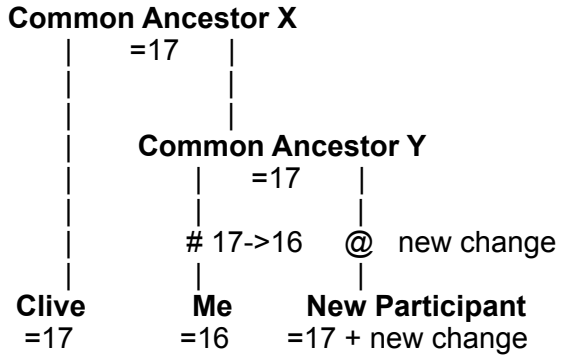
Another line of research is to look for matches beyond the Warburton surname. There are websites where profiles can be stored and matched. It is possible a match may be close enough to determine the surname of the father in a non-paternal event. However it is also possible that such a match is merely random so other evidence is needed to prove a match. I have found a number of matches with non-Warburtons, a couple of which have additional factors to encourage the belief they are genuine matches. These are discussed in the **Commentary on Results**.

Phylogenetic Trees

As well as directing our more traditional genealogical research, the DNA study can employ some new techniques. When there are a few matching results they can be linked together in a phylogenetic tree, which is a simplified family tree in which the location of changes is deduced as far as possible.

As an example, the one difference between my genetic cousin Clive and myself is that he has 17 repeats at marker DYS458 and I have 16. Subsequent results have also had a value of 17, indicating that 17 is probably the original value and the mutation to 16 took place somewhere in my line.

If we tested a new participant with whom I shared a more recent common ancestor, we could pinpoint where the change occurred more accurately, based on whether he had a value of 16, or 17. Furthermore, if the new participant had a different change from me, any future clan member showing the same change would be closely related to him. See the example genetic family tree below.



=17 or =16 the value of DYS458
position of change from 17 to 16 (roughly: it could be anywhere between common ancestor Y and Me)
@ rough position of new change

I have produced a **Phylogenetic Tree** for the 30% of results that come from the North Cheshire clans.

Handling Grey Area Matches

Only a couple of results have shown a low probability of a common ancestor in 25 generations. In such cases intermediate results might increase confidence in a match. For example they might both be more closely matched to a third result.

There is another way of evaluating these low probability matches, and this is to consider the probability of a random match, particularly where the participants share a low frequency allele for certain markers.

The probability of 2 random individuals having the same values for a set of markers can be calculated using the allele distributions. For example if the probability of a specific value at one marker is 90% and the possibility of another specific value at another marker is 80% then the probability that 2 people have the same values is 72% (0.9 x 0.8 expressed as a percentage). If you multiply the probabilities of having the most common allele for each of the 32 markers common to **DNA Heritage** and **Family Tree DNA**, then the probability of 2 random individuals having a complete match is just 6 one thousandths of one percent. The chances of sharing some of the less common alleles are even less, particularly when there are more than 2 participants in the potential group.

We must also acknowledge the possibility that a common ancestor could be a pre-Warburton, or a non-Warburton.The village of Warburton was founded in the 10th century as a fortified settlement near a ford on the river Mersey. I would imagine it was established with a handful of families, and 3-400 years later when surnames were being adopted it would still be largely populated by descendants of these families. This would be particularly true of the males, although wives might come from neighbouring villages. It is possible, therefore, that two men from the village who were already related might independently adopt the Warburton name. We therefore have the possibility of a pre-Warburton common ancestor up to 1050 years, or 35 generations ago. As an aside it is worth commenting that while Sir Piers de Werberton adopted the Warburton name when he moved to Warburton, as it identified him in regional or national circles, an adopter from the lower classes would probably only take the name when he moved away, to signify his origin.

The other possibility is a non-Warburton common ancestor. I have noticed it is not unusual for there to be multiple marriages between two families. It must be possible, therefore, that there 2 Warburton girls, maybe from different generations, have illegitimate sons by members of the same local family, thus giving them a common non-Warburton ancestor.

Summary

DNA testing is best viewed as an additional tool in traditional genealogy. The various types of DNA test can add to your understanding of where you come from. In particular Y-chromosome DNA profiles can be extremely useful in the study of family history if they are used in addition to the traditional tools of genealogical research. My objectives in the **Warburton One-Name Stud**y is to continue to exploit this synergy to develop an ever broader understanding of the various Warburton clans and what links may exist between them.

References

1. Stephen Oppenheimer, *Out of Eden: The Peopling of the World* (Robinson Publishing, 29 Jul 2004)

2. Professor Bryan Sykes, *Seven Daughters of Eve* (Corgi, 1 Sep 2004)

3. Professor Bryan Sykes, *Adam's Curse: A Future Without Men* (Corgi, 1 Sep 2004)

4. Chris Pomery and Steve Jones, *DNA and Family History: How Genetic Testing Can Advance Your Genealogical Research* (PRO Publications. Sep 2004)

5. Professor Bryan Sykes, *Blood of the Isles* (Corgi, 3 Sep 2007)

6. Stephen Oppenheimer, *The Origins of the British: A Genetic Detective Story* (Robinson Publishing, 12 Apr 2007)