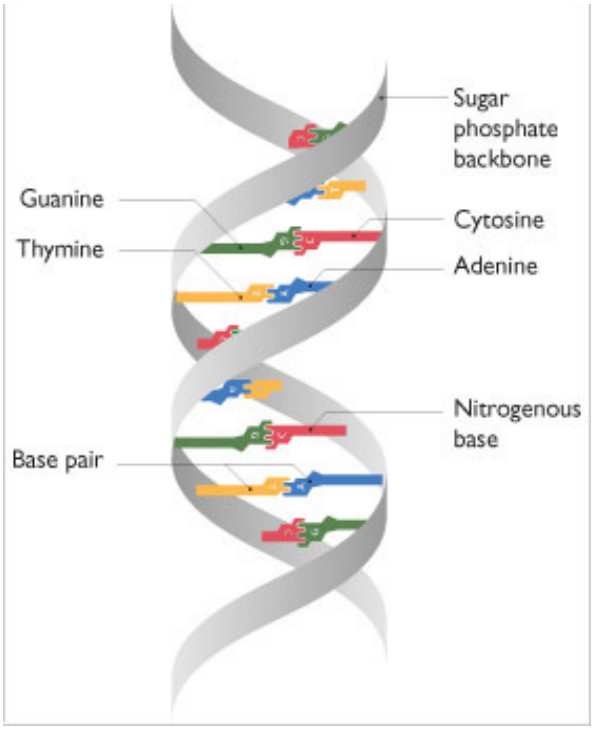
DNA and its Uses in Genealogy

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. At the end there is a reading list of books I have read 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 body of the cell (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 set 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 three categories of DNA test marketed for people interested in their genealogy or genetic history. These are Y-chromosome tests, mitochondrial tests and autosomal tests. **The Warburton DNA Project** focusses on Y-chromosome tests because, like a surname, they pass from father to son, and so fit neatly with a one-name study. I will discuss mitochondrial and autosomal tests briefly at the end.

Y-Chromosome Tests for Genealogy

The Y-chromosome tests used in genealogy are Single Nucleotide Polymorphism (SNP) tests and Short Tandem Repeat (STR) tests. Full Sequencing SNP tests were originally extremely expensive, therefore attention was focussed on the STR tests. However prices have dropped considerably, so although still expensive when compared with STR tests, SNP testing is now practical for genealogy testing in many circumstances. STR tests do still have a useful but diminishing role.

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. So a SNP is when a letter is copied wrongly.

SNPs 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 SNP and those who don't. Of those who do, a proportion will have a later SNP, and so on. All the people who share a particular SNP form a haplogroup. Haplotrees are charts that map the relationships between haplogroups. By relating the current geographical spread of haplogroups to archaeological evidence, and the amount of change at various locations, researchers have been able to piece together ancient human migration patterns.

Family Tree DNA have developed a Discover Tool that encompasses a family tree of humankind where all male lineages trace back to a single common ancestor who lived hundreds of thousands of years ago. This human tree allows you to explore lineages through time and place and to uncover the modern history of your direct paternal surname line and the ancient history of our shared ancestors.

SNPs are discovered by a next-generation sequencing test. This will trawl through a large part of the Y-chromosome checking your DNA against all the known SNPs in that part of the Y-chromosome, and looking for new ones unique to you. There are several suppliers of these tests, testing greater or lesser parts of the Y-chromosome. Obviously the more they test the higher the cost.

It is also possible to test for specific known SNPs, either singly or in groups. However without some prior indication of the SNPs you are likely to have this is a hit and miss approach akin to searching for a needle in a haystack. It might be useful if your match has determined his SNPs in detail and you only wish to know which of his more recent SNPs you share.

The **Family Tree DNA** full sequencing test is the Big Y-700 . It covers up to 23 million base pairs of the Y-chromosome, and also includes over 700 STR markers. The Big Y-700 results include a diagram showing your position on the Big Y Block Tree, along with matches, neighbours, and the sequence of SNPs which lead there. It has been calculated that SNPs uncovered by the Big Y-700 test, occur on average, every 83 years, giving a method for calculating the time to the most recent common ancestor.

It was by taking this test that I discovered my most recent shared SNP is R-FGC17094, and I have more recent ones unique to me. SNP R-FGC17094 occurred since the Warburton name was adopted, and is probably about 500 years old. Thus anyone sharing that SNP is a Warburton (or descended from one) and shares a common ancestor with me.

Short Tandem Repeat Tests

When Y-chromosome testing for genealogy began in earnest, comprehensive SNP testing was expensive. Fortunately there was another test that was much cheaper, and can be used to determine matches that denote a common ancestor, particularly if supported by a shared surname, or other historical evidence.

It so happens that there are some short DNA sequences that are repeated several times. Whereas with SNPs 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 where they occur on the Y-chromosome. These are called markers in STR tests. There are tests available for different numbers of markers, and different testing companies test different markers.When two STR results are compared the number of markers with different values are counted. This number is known as the genetic difference.

Time to Most Recent Common Ancestor (TMRCA) calculations use genetic distance, in conjunction with mutation rates for individual markers, to calculate the number of generations since a common ancestor lived. I did a lot of such calculations early in the project, but now the emphasis is much more on SNP testing and the **Discover Tool**. I have therefore put detailed discussion of the techniques into a separate document called STR **Time to Most Recent Common Ancestor Calculations.**

At its inception the **Warburton DNA Project** was based on STR tests, beginning with 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 37, 67, and 111 markers. The **Y**37 STR test provides sufficient markers to be able to determine matches in a surname study.

My DNA Results

My BigY-700 test result showed I have SNP U106 which is the subject of a major project at **Family Tree DNA**, and SNP R-DF98 which identifies me as belonging to a group called **The King's Cluster** within the R-U106 project.

My direct ancestors were in modern Western Germany about 5,000 years ago. They disappeared a thousand years later only to re-emerge in Britain following the Norman conquest. As there has been little DNA testing in France I presume this is where they spent the missing period.

My more recent SNPs confirm an earlier STR match with a group of Duttons. History records that it was a descendant of a Norman knight, Odard de Dutton, who first adopted the Warburton name after building a manor house in the village of the same name.

The difference between STR matches and SNP matches is that SNP matches are definitive whilst STR matches are only indicative. My most recent shared SNP, FGC17094, occurred once in a particular individual. My match John and I are his descendants.

FGC17094 is at the end of a long list of SNPs which define our history. Others, like Mark Warburton, the Duttons and the Howells separate from this line at earlier stages. Another cousin, Clive did a specific SNP test for FGC17094, and was positive, proving he shared our history.

John and I had an earlier STR match. This indicated we had a common ancestor, but because STR mutations are bi-directional it was only the additional fact of our common surname that verified this. The test only indicated our earliest SNPs. The dates of common ancestors can be estimated from the genetic distance between two STR tests, and patterns of mutations can give some idea of when related clans diverged, but SNP sequences and matches give a more precise picture.

Unmatched Results

To date the success rate in finding matches is over 70%. This leaves under 30% of results with no match. There could be many reasons for this, We don't know how many people just adopted the name in feudal times. Many lines will have died out, and others could have few descendants today, even though they are old. Other lines will have started later as a result of adoption, illegitimacy, or infidelity, but may still have had time to grow into significant clans. Bancroft Warburton was illegitimate when he was born in 1738 (though his parents did later marry) but his descendants include over 170 Warburtons.

These later events are referred to as non-paternal events. Known non-paternal events account for nearly half of the unmatched Warburton results. Whereas DNA profiles always pass from father to son, 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.

Even a small number on 'non-paternal events' in each generation will add up over several generations and many of us will be affected by them somewhere in our family tree, if not in our paternal line. We all have 32 great great great grandparents. Add in parents, grandparents etc. and there are 62 individuals we are directly descended from in a tree that goes back just 200 years. The law of averages says at least one of these will be the result of a 'non-paternal event'. I know I have one illegitimate great grandmother, and a great great grandmother I'm not sure about.

I have also uncovered, and written about, a case in the Warburton Village clan where matches show that early 18th century two cousins have different DNA profiles. This means a non-paternal event must have occurred, but there is no evidence of it in the parish record.

Of course by definition every Warburton line must have started with a 'non-paternal event'. The first Warburton in the line must have had a father whose name wasn't Warburton. 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 in many cases.

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.

Unmatched profiles could still belong to sizeable clans and have ancient origins so there is a chance they will be matched in the future. I've found a couple of new matches in the past year.

There is also the possibility of identifying links to other surnames and uncovering clues as to your true biological origins. For example I have one project member who discovered he had matches with a number of Stewarts. With the advent of the Y-Tree and the Big Y Block Tree Matching Tool such matches will become much easier to find because you can be sure SNP matches are real and not possibly just random results.

Triangulation

When the results of two participants in the **Warburton DNA Project** match, this identifies a triangle with the participants' common ancestor is at the apex, and the participants at the base. All the other male Warburton descendants of the common ancestor are contained within this triangle, and will have the same DNA profile as the first two participants, unless of course there was a '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. The only way to determine at least a minimum time of association is to find a matching profile, and build a triangle.

Where a match is found and a triangle created, this triangle can be fitted onto the Warburton clan family trees. A triangle may only cover part of a clan, with its common ancestor being a descendant of the clan's earliest ancestor. We would not be certain at this stage if we had found the DNA profile of the whole clan. I have one clan, the Warburton Village clan, which contains two triangles with different profiles, and where the common ancestors of the triangles are cousins who lived in the early 18ty century. At least one of these profiles must result from a 'non-paternal event' though nothing untoward is documented in the parish records.

When the match is with a member of a different clan, 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 ancestors are 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 known 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. 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.

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 haplotree (see below), but this could be offset by the risk of uncovering an unknown, recent 'non-paternal event'.

Mitochondrial Tests

Mitochondrial tests are SNP tests. As mentioned above, Mitochondria is a small piece of DNA that acts a the energy source for a cell. It has roughly 16,000 bases, which means SNPs occur much less frequently than for the Y-chromosome. However, like the Y-chromosome, because it only passes down the female line it remains unchanged from generation to generation.

As we all carry mitochondria inherited from our mothers, anyone can take the mitochondrial SNP test. The small number of SNPs limits its use as a genealogical tool, though it was the test that identified a skeleton found under a Leicester car park as that of King Richard III.

It does provide an alternative view of our deep history. 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 in a companion paper **My Use of DNA in Genealogy**.

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,** and **AncestryDNA** from **Ancestry,** test DNA from the 22 pairs of chromosomes that contain a mix of DNA from all ancestors. Strings of DNA are compared either between databases of 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 fifth cousins. These tests are also relatively cheap leading to them being the most popular tests used today. In January 2019 it is estimated that 15 million people have tested so far, 10 million of these using the **AncestryDNA** test.

The value from these tests come from the matching databases maintained by the suppliers. There is also a cross supplier matching database at gedmatch.com, but it seems that only about 1 million results are loaded there. Matching is a more technical exercise than for SNP or STR matches. A SNP match is either positive or negative, and it is straightforward to compare two STR results and calculate genetic difference. Autosomal comparisons involve scanning large amounts of DNA from databases of results looking for strings of DNA that are common, and this can only be done by the owners of the databases.

Like all DNA tests autosomal tests face the possibility of an unwelcome result, possibly exacerbated by the fact that they concentrate on recent generations. If a Y-chromosome test indicates non-paternal event it could have occurred hundreds of years ago, unless you test two closely related males, and this isn't usually necessary. However an autosomal test can uncover recent family skeletons. A new meaning of NPE is used; not parent expected.

Some early tests reported small amounts of Neanderthal and other pre-human DNA in most non-Africans. 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.

However I have noticed that results from the currently most-popular testing companies no longer show this element of our DNA.

Summary

DNA testing is best viewed as an important 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.

Further Reading

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)