# The Genes in Genealogy

The Scientific Basis of Heredity

## Recent Historical Background.

Watson and Crick reported the structure of the DNA molecule in the journal Nature in 1953. Thirty years later, in 1983, invention of the polymerase chain reaction (PCR) for rapidly copying DNA from small samples made it practical to do large numbers of tests. Workers in human anthropology began to use this information to produce "histories" of human migrations and the broad relationships among the various streams of humanity across the globe. In June of 2000, President Clinton announced that the mapping of the human genome was essentially complete. Each (actually, about 90%) of the "letters" of the molecular instructions for making people had been read. Since then changes in technology, costs, and the market made it profitable to offer DNA testing to individuals interested in personal genealogy. Appropriate use of this new opportunity requires understanding of some of the background scientific knowledge being applied. Genealogists may find an interest in this new area of research and, in a few cases, may find solutions to special problems in standard genealogical work. The field is still developing and we may expect continuing changes in the tests, interpretations and costs

## Our Genetic Inheritance.

Our bodies are built according to chemical instructions recorded in our inherited genetic material. The information necessary to carry out the complex chain of chemical events which creates our bodies is encoded by the sequence of components which make up the DNA material received from our parents. The structure of the DNA molecule is a wonder, both in form and function. As is now well known, the DNA molecule is built in the form of a twisted ladder. The sides of the ladder are connected segments containing sugar and phosphate building blocks. The rungs of the ladder are loosely bonded pairs of nucleotide molecules. In fact, only four nucleotides are used in DNA. They differ only in one part, containing the organic bases, which are adenine, cytosine, guanine and thymine, aptly abbreviated A, C, G, and T. Since A only bonds with T and C only with G, any one rung of the ladder is either A==T or C==G. At first glance, these seem to be very minimal variations to distinguish between you and me or between me and a turnip! The required variations arise in the order in which these rungs are arranged. In a further simplification of our notation, only one side of the chain of ladder rungs need be reported because an A only links with a T and a C only with a G so we can tell what both sides must be while only looking at one. A short segment of my genetic code can therefore be reported as ATTCTAATTT. In a long list, there are many possibilities for different sequences which then produce the many forms of life, including more than seven billion unique human beings.

As cells divide to form new cells, each chromosome (DNA molecule) "unzips" along the links between the nucleotides (an A comes apart from its neighboring T, for example). Then each side of the ladder builds a copy of its former partner from the molecular soup in which it floats. An A can only bond with T and a C with a G, as before. The newly built half-ladder can only be constructed in the way which replicates the lost partner. Each chromosome is now represented by two copies of itself, one of which manages to move aside into each of the about-to-be created new cells. This is the absolutely amazing way in which each cell passes on its exact genetic code to its descendants.

Our DNA is contained in 23 pairs of chromosomes in the nuclei of our cells plus a small amount of DNA in the mitochondria outside the nuclei. These structures contain a little more than three billion "base pairs" (like A==T) which code the necessary information. All humans are identical in 99.9% of this code, much of which we also hold in common with our relatives in the animal kingdom. Among the closest of these are the chimpanzees and gorillas. The remaining 0.1% of the code, which accounts for the variation among us, contains about three million base pairs.

When germ cells are created, cell division follows a different process called meiosis. The chromosomes do not replicate, but they line up two by two and a new cell well forms between the line of pairs. In this way, the germ cell contains only half of the DNA of the parent cell. When sperm and egg are united at fertilization, the new individual has the full complement of chromosomes and one member of each pair comes from one parent with the matching chromosome coming from the other parent. In this way, we get half of our DNA (one chromosome of each pair) from each parent. Our nonidentical siblings, however, will not get exactly the same set of chromosomes from each parent so on the average we share only 50% of our DNA with a sibling..

Although we inherit half of our genetic material from each parent, we do not generally get an equal amount from each of our ancestors in a more remote generation (e. g., the chromosome of pair 1 which I got from my mother may have come from either her father or her mother, but not both). We have 64 fourth-great grandparents and only 46 chromosomes. Somebody gets left out! Our genetic inheritance is not quite that simple, however. By a process known as recombination, chromosomes within a given pair may exchange a portion of their genetic material at the time of creation of egg or sperm cells. The process of recombination provides for the great variety of human genotypes which allows each of us to have a unique mixture of genetic material from our ancestors – unless, by the natural cloning process, we happen to have an identical twin or triplet. While recombination gives us each a special genetic code, it also makes it difficult to say which portion, if any, of our genetic material came from a particular remote ancestor. I can't say which remote ancestor a specific bit of my recombinant DNA comes from nor can I predict which other living descendants might share it.

Certain portions of our DNA are not subject to recombination. One such portion is the mitochondrial DNA, abbreviated as mtDNA, which is not in the nucleus, does not undergo

meiosis nor recombination, and is passed to us nearly unchanged from our mothers, their mothers, etc. back through time. When an egg cell is fertilized, it actively excludes or kills the mitochondria from the sperm cell. Another portion of DNA not subject to recombination resides on the Y-chromosome. Our 23rd pair of chromosomes is not really a pair at all. They are the X and Y chromosomes. The Y-chromosome is a stunted little thing, carrying relatively little genetic material. It is not a physical match for the related X chromosome and cannot, for the most part, participate in recombination. The Y-chromosome does have one special feature. It makes males rather than females. Only males have the Y-chromosome and a child inheriting the Y-chromosome from its father will be a male child. This genetic material, in a male, will have been inherited nearly unchanged from the father, the father's father, etc. back through time.

The reason the genetic code is not identical down to the very last letter from one human to another is that mutations have occurred. While the DNA strands separate and replicate, an occasional "copying error" is made. If this happened a lot, we would all be completely different from our parents (and probably all dead because most errors would not be viable). If it never happened, we would all be genetically alike.

## Genealogical Use.

**Mitochondrial tests**. Scientists, hoping to distinguish among human beings and to relate them to each other, have become interested in only a small part of that 0.1% of the human genome which differs from person to person. Any portion that is the same for all people will not help in distinguishing them. On the other hand, if any portion is randomly different for all people, it will not help in discerning relationships. Further, if any part of my own genetic code may have come from any one of my 64 fourth great-grandparents, it cannot be used to define a specific ancestral line. Only that portion of the 0.1% of the code that varies among humans and is non-recombinant is currently useful in making statements about specific lines of descent.

Anthropologists, studying the development of the human race, needed to find portions of the genetic code that distinguished one human from another but did not mutate so rapidly that no firm connections could be made across time. Certain locations in the mitochondrial DNA ring were found to be useful in distinguishing major segments of the human population. Changes which happened only every ten thousand years or so were helpful in classifying major population groups and, coupled with maps of the current geographic distributions of the resulting mutations, supported discussions of the migratory history of those groups. These mutations were not only non-lethal, they, for the most part, have no known function and appear to be in "nonsense syllables" in our genetic code.

When chronological trees of these mutations were studied and statistical clusters of the changes were created, Dr. Bryan Sykes found that the present population of Europe is descended from only seven distinct women (The Seven Daughters of Eve). His characterization and personification of these seven women (he gave each a name and a life story) have popularized the field and, I think, created the current market for genetic testing for personal interest. These

women lived at very different times, in very different cultures, but each just happened to have daughters who had daughters who had daughters etc. who eventually were the mothers of nearly all the living people of European descent who were tesed. Although several other such "Daughters of Eve" are found in worldwide studies, the small number of them captures the imagination. As a classical genealogist, however, it is of little value to me to learn that I am a descendant of "Katrine" (i. e. haplogroup K) who lived 15,000 years ago even if I learn that one or two other people in this room share this descent, making us something like 650th cousins. Nevertheless, the question is no longer "Are we related?" but "When did our most recent common ancestor live?".

The mitochondrial DNA molecule is a ring containing just over 16,500 base pairs. The regions initially used for our purposes are called HVR1 and HVR2. HVR means hyper-variable region. It changes often enough to be useful on a time scale of tens of thousands of years. One lab I know tests 400 letters of code in HVR1 and another tests 540 letters in HVR1 and 570 letters in HVR2. They report your code and where it differs from a standard reference code (not from the code of Mitochondrial Eve who has not been tested). More recent tests now report on the entire mitochondrial ring.

**Y-chromosome tests**. All people have mitochondrial DNA available for testing. The other nonrecombinant DNA resides on the Y-chromosome which is only carried by men. This chromosome is inherited from the father and thus follows what we might call the "surname line" except that the surname is historically limited and culturally derived. Women may have their "surname line" tested by enlisting the aid of an appropriate male relative carrying the relevant Y-chromosome. Apparently, single letter changes (Single Nucleotide Polymorphisms, SNPs) occur even more slowly in Y-DNA than in mtDNA and it was difficult to find proper markers to distinguish groups of people. In fact, initial tests showed no difference at all. However, anthropologists eventually found regions of the Y-chromosome, apparently useful for nothing else, which vary at about the rate required (not too fast, not too slow, but "just right"). In fact, SNP mutations are so rare that each is assumed to have happened only once in the human race. This implies that, if I have the mutation, I descend from the first man who had it. The genealogy of these mutations has been mapped into a family tree onto which I could place myself but it groups humanity into such major groups that it tells little about my family in a genealogical time frame. These haplogroups are based on SNPs as were the mitochondrial daughters of Eve. There are maps of the distributions of these haplogroups across the world.

A group of Y-chromosome locations which changed more rapidly was needed in order to make finer distinctions. Labs now test twelve, twenty-five, or more such locations. However, the contents of these locations change so often that each mutation cannot be considered to have occurred only once in the history of mankind and neither can they be arranged into a chronological order nor family tree sequence because some change back again in a later generation. They do, however, change on a time scale of hundreds of years, which might be relevant to genealogy within historical times.

These more frequent DNA changes are of a different type than finding a location where a T has become a C, however. The locations of interest show multiple repetitions of a short sequence of base–pair codes which are called Short Tandem Repeats or STRs. It is as if there is a "stutter" in the coding process. At each tested location, there is a specific number of repeats of a given coded sequence. The number of repeats in a given segment is counted, reported and used as a distinguishing factor between individuals and families. A report might, for example, show DYS#390,11 which, being translated, means DNA Y-Chromosome Sequence 390 has a repeat count of 11. Another person's report might show a count of 13 at that location. When you get your report, you can go on-line and learn the frequencies, in various samples of humanity, of the various alleles at your test locations and attempt to locate people who match or nearly match you in each, or most, locations. There is much discussion of various statistical conclusions based on matching tests. For example, it is claimed that if two people match at 37 out of 37 tested locations, there is at 50% probability that their Most Recent Common Ancestor (MRCA) was within the last five generations and a 95% probability that he was within the last 8 generations.

Although Y haplogroups recognized by anthropologists are based on SNPs, enough statistical correlations have been made to allow the Y haplotypes based on STRs to be placed in the haplogroups. Thus, you can learn your haplogroup without SNP testing. Mine is R1b which is also the most commonly found group in the US and is predominant in western Europe, thought to be descendants of men who arrived there before the last Ice Age about 40,000 years ago. It was not completely satisfying to have this non-definitive haplogroup assignment, so I had further testing done in SNPs and I am currently classified in haplogroup R-Z18 in a newer classification nomenclature.

**Autosomal tests:** Some useful testing can be performed on the recombinant DNA as well. One test uses locations found to vary significantly among various ethnic groups. By making tests which include several SNP locations across many chromosomes and correlating the results with individuals of known ethnic heritage, statistical statements can be made about the subject's own heritage. These results are labeled "Recent Ethnic Origin" because, as we have seen, more remote ancestors may not be well represented in the recombinant DNA of living descendants. Statistically, each of our 64 fourth great-grandparents would give us about 1.6% of our DNA and this is below the reported margin of error of the testing. So, unless that elusive "pure" American Indian in your ancestry is rather recent, he or she may not be detected by these tests. There are also special Native American tests for the male and female lines. Remember that some such tests would examine only one male and one female ancestor in each remote generation (not very much of your whole ancestral chart).

Additional tests, using different portions of the DNA, are used for forensic purposes or for medical purposes. Most labs were careful to state that your genealogical tests will not provide you or anyone else data that might relate to your medical problems or, at least, that such links are not known. This did not remain true in the face of further knowledge. One lab compared your own results against data bases collected for cancer (or other disease) research and could provide reports on your statistical tendencies to show a variety of genetically influenced conditions. The

FDA recently ordered them to discontinue this type of reporting.

Autosomal tests for genealogical purposes, however, remain popular. Scientists have identified hundreds of thousands of autosomal SNPs which prove helpful in indication relationships among people. After obtaining your results from an autosomal test you may post those results on GEDMATCH.COM and a computer program will compare your results with all others posted there and show you where and on what chromosome you match with others. These are potential relatives. FamilyTreeDNA and Ancestry.com will also report matches you have with others among their customers who have shared their results.

Another result from autosomal testing comes in the form of a chart or map showing what percentage of your ancestry comes from what region of the world. As the database of DNA results grows and the software used for analysis becomes more robust, such maps become more specific. Although I cannot find any support for the timing of this map, I have read that it may represent the regions where my ancestors were roughly 1000 years ago. [Of course, my DNA was *somewhere* at any time in human history. DO MORE WORK ON THIS.]

## SUGGESTIONS AND PROCESS.

I would recommend that you first read some more about DNA inheritance and the uses of DNA testing in genealogy. Many popular articles are now available. When you understand more about the process, you will be better able to judge whether some of the tests will be of help or interest to you. You may have to apply your genealogical and recruiting skills to locate people whose DNA will be relevant to your problem and to convince them to help you. Good design requires that you have selected a genealogical problem, the right test for its solution, and the right group of participants. Or, you could do what I did and let your curiosity overcome any need for design.

There are several success stories already. Several families now have DNA evidence to support traditions that two groups, with the same or similar surnames, do have a genetic connection. In other cases, traditions that the surname used to be something else can be supported by evidence that the two families of different surnames share the same Y-DNA. mtDNA studies have been helpful in allocating the several children of one man among his two or more wives.

In my case, another man started a surname project for the OLD family. He recruited a few male-line descendants of the OLD family in Virginia and showed they all have a common origin, even though the paper trail does not show exactly who the common ancestor is. When I compared my own results with those in this project, I found that I differ from them by only one count in one marker of the twelve they tested (now 3/37). Thus, the OLDS family of New England probably has a common origin with the OLD family of Virginia (both families actually use both spellings). This supports the tradition that both families derive from the same family in Dorset, England. It does not prove who or where our common ancestor was.

Read more about genealogical testing companies (perhaps online) to see if any have tests available that would help you and to compare services and prices. If you order a test kit, the usual process will be for you to use a swab in the kit to collect some of the loose cells from the inside cheek wall in your mouth – a simple, painless procedure. Mail in the test kit and await your results.

The lab will extract a sample of DNA from the cells you send and "amplify" selected segments of it. The variable DNA segments of interest are identified by locating the non-changing region which neighbors them and then cutting out the appropriate portion for further study. Special "designer chemicals" are used that search out and extract unique segments of the code to be tested and a repeated thermal cycling process uses the inherent ability of DNA to replicate itself under the right conditions. If DNA segments are doubled twenty times by such a repetitive process, each original DNA segment then has more than a million copies ready for testing.

The resultant copies are caused, by an electric field, to move through a special gel material. The distance they move in a given time depends on the length of the fragment. By using known fragment lengths for a calibration, the unknown lengths can be determined. Finally, a report can be made saying something like: At Y-DNA location 390, your DNA repeats the code GATA eleven times. Measurements of several such locations give the desired results. Sequencing the mitochondrial DNA for SNPs is a slightly different process. In the "soup" of materials from which the DNA is being constructed, specially modified A, C, T, and G molecules have been placed. These molecules have portions which will fluoresce (a unique color for each type) and have also been modified at one end so they cannot link to the next member and the chain ends. When the chains are now sorted by length, some will found, for example, to have length 11 and fluoresce the color associated with A. Your DNA has an A at that location in its structure. Do enough repeats, gather results for all lengths, and you can deduce the code for the whole DNA segment. The final report will show something like: Your mitochondrial DNA at location 16311 has a C rather than the T in the "standard" code.

When you have your results, compare your DNA with samples now posted on several websites for the purpose, probably including one managed by your testing company. Post your own results so that others may contact you about possible relationships. Learn how others interpret their results for your haplogroup by reading further on these websites or by participating in DNA and surname related mailing lists. Answering the question "What do these numbers mean" requires more than a single lecture – perhaps more that a single course. The answer will be in statistical terms such as "These two men have a 50% likelihood of having a common male-line ancestor within the last seven generations." This can be useful to anthropologists but statistics is not genealogy, in my opinion. These statistical results **can** guide further "classical" genealogical research. Results which indicate very low probability of recent common ancestry can be more conclusive and, though perhaps disappointing, are also helpful. The real meaning for genealogy will be in the matches with other people that are, or are not, established when results are compared.

## CONCLUSION.

While it is of interest to read conclusions about the development and migrations of the human race, this is not genealogy but anthropology. The existence of a mitochondrial Eve and a Y-Adam from whom we all descend but who lived many tens of thousands of years apart is now well accepted although the suggested dates are still changing. Computer models show – based on assumptions about human movement between populations – that all humans alive today share some unknown and unknowable common ancestor who lived only a few thousand years ago. Within a thousand or so years before that, all humans then living are ancestors of ALL people living today or of NONE. Naturally, there is some healthy doubt about the accuracy of such models.

Certain DNA studies can now be used to supplement or guide traditional **genealogical** studies. They can confirm or (especially) deny a family connection but will not identify a specific common ancestor because the same DNA would normally also be found among the ancestor's other relatives as well. Regardless of results, this new technique has provided me with the incentive to embark upon a fascinating new learning adventure.

## DEFINITIONS.

Allele – one of a number of alternative forms that can occupy a given location on a chromosome.

Base – one of the four elements of the genetic code – Adenine, Cytosine, Guanine, and Thymine, commonly abbreviated A, C, G, and T.

Chromosome – a thread-like structure in the cell nucleus; contains DNA. Chromosomes are visible in microscopic studies of the nucleus of the human cell.

DNA – Deoxyribonucleic acid, the molecule which contains the genetic code.

Gene – segments of the genetic code which have specific results in development of the individual. There are perhaps 30,000 human genes which comprise only about 2% of the human genome.

Genome – the entire genetic code for an individual. The human genome contains more than three billion base pairs. The "full sequencing" of the human genome was announced in April 2003.

Haplogroup – a statistical grouping of similar haplotypes

Haplotype – a classification of people based on a set of results from genetic testing. Which

"set of results" is used varies with the context in which it is used and some study may be required to determine the meaning the author intends.

Homo Sapiens Sapiens — our variety of our species, "modern man," appeared more than 150,000 years ago in Africa. Estimated ages of this nature vary considerably. It is possible that Mitochondrial Eve is older than Homo Sapiens Sapiens.

Meiosis – cell division (mitosis) which results in gamete (sperm or egg) production

Mitochondria – organelles in the cytoplasm (non-nuclear material) of our cells which help in the utilization of energy by the cell.

Mitochondrial DNA (mtDNA) – DNA which resides in the mitochondria and is inherited only from the mother. We have about 16,500 base pairs in our mitochondrial DNA.

Mitochondrial Eve – proposed in 1983 as the most recent female-line ancestor of all living humans, now generally accepted as having lived in Africa perhaps 140,000 to 200,000 years ago. Of course many other women were alive at that time also but the maternal line of descent from each of others has died out.

Mitosis — division of the nucleus (or cell) which results in two duplicate nuclei (or cells)

Mutation – a change in the genetic code caused by a copying error in transmission from a cell to its descendants. An average mutation rate for the Y-chromosome STR markers used in genealogy is .002 or one in five hundred transmission events.

Non-recombinant DNA – a portion of the individual's DNA which does not undergo recombination.

Nucleotide – a building block of the DNA molecule containing a sugar/phosphate group and one of the four bases.

Polymerase Chain Reaction – a reaction that emulates the natural DNA replication process and facilitates the production of millions of copies of selected segments of a DNA sample.

Polymorphic – or "exhibiting more than one form." Used to refer to genetic locations which may differ from person to person.

Recombination – the process by which some portions of the DNA molecule are exchanged between paired chromosomes during meiosis

Single Nucleotide Polymorphism (SNP, pronounced "snip") – a genetic code difference at one specific location (i.e. in only one "letter" of the code).

STR – Short Tandem Repeat. A short sequence of genetic code which is repeated several times in sequence. Y-chromosome STRs are used in testing for family relationships by counting the number of repeats found at each of several locations.

X chromosome – This chromosome exists in the 23rd chromosomal pair in all humans. If it is paired with another X, the person is a female; if paired with a Y-chromosome, the person is a male.

Y-Adam – The most recent man who is the paternal line common ancestor of every living human. He lived in Africa about 90 thousand years ago. Estimates of age do vary greatly.

Y chromosome – The chromosome that exits only in males. It is small, containing only about 60 million base pairs. Many extended studies found very few or no SNP mutations in the Y-chromosome. It had been previously noticed by microscopic examination, even before DNA sequencing, that Y-chromosomes were not all the same length. This has lead to the use of counting repeats (STRs) of segments of DNA as a means of distinguishing one group of men from another. SNPs have been found and are now used as well..

# Some Reading

Printed resources


Aulicino, Emily D., Genetic Genealogy, The Basics and Beyond (published by AuthorHouse, 2104). This is the most recent book I have used and therefore the best on recent topics such as autosomal testing.

Pomeroy, Chris, DNA and Family History: How Genetic Testing can Advance your Genealogical Research (published by Pro Publications, 2004)

Savin, Alan, DNA for Family Historians (published by Alan Savin, Maidenhead, ENG. Alan@savin.org, 2002)

Smolenyak, Megan Smolenyak and Turner, Ann, Trace Your Roots with DNA, Using Genetic Tests to Explore Your Family Tree (published by Rodale and distributed by Holtzbrink Publishers, 2004)

Sykes, Bryan, The Seven Daughters of Eve (W. W. Norton, NY & London, 2001). A book which is relatively easy to read in which Dr. Sykes of Oxford popularized the anthropological research into DNA and in which he gave names, stories and personalities to each of the seven sources of modern European DNA. This popularization seems to be the driving force that created the current interest in DNA testing for genealogists which has spawned several business enterprises, including Dr. Sykes' own Oxford Ancestors.

Wells, Spencer, The Journey of Man, A Genetic Odyssey (published by Random House, New York, 2002). This is anthropology, not genealogy, but it uses the same DNA we do and, in fact, the anthropological uses predate the genealogical ones.


Internet resources (tested and revised., June 2014)

http://www.kerchner.com/dna-info.htm – a good collection of explanations and further references.

http://www.scientific.org/tutorials/articles/riley/riley.html An introductory discussion for non-scientists of the DNA testing process.

http://nitro.biosci.arizona.edu/ftdna/TMRCA.html – a calculator for finding the time to the Most Recent Common Ancestor, given genetic test results (Y-DNA) for two individuals [now old]

http://www.cyndislist.com/dna.htm The relevant entry in the most famous genealogical reference list of internet resources.

www.isogg.org/wiki/Autosomal_DNA_Testing_Comparison_Chart
A detailed comparison of autosomal DNA testing offers from five different companies. There are many other good articles in this Wiki. www.isogg.org is the site for the International Society of Genetic Genealogists. Anyone can join and they want volunteers of time and money, but nothing is required.

www.dna-explained.com/2012/08/14/y-dna-family-tree-dna-vs-ancestry
A careful comparison of two Y-DNA tests

GENEALOGY-DNA-L

*Topic:* This mailing list is for anyone with DNA (i.e., anyone!) who would like to discuss methods and share results of DNA testing as applied to genealogical research.

For questions about this list, contact the list administrator at GENEALOGY-DNA-admin@rootsweb.com.

*Subscribing:* to join **GENEALOGY-DNA-L**, send mail to GENEALOGY-DNA-L-request@rootsweb.com with the single word *subscribe* in the message subject and body. To join **GENEALOGY-DNA-D,** do the same thing with GENEALOGY-DNA-D-request@rootsweb.com. This is the digest form of the list which sends less frequent e-mail compilations rather than each individual posting.

Some testing services – the websites of these business all contain useful educational material.

www.23andMe.com – until recently 23andMe's marketing edge was defined by their reports on the health connections of the DNA results. They are now attempting to negotiate through and FDA edict that they stop providing that analysis.

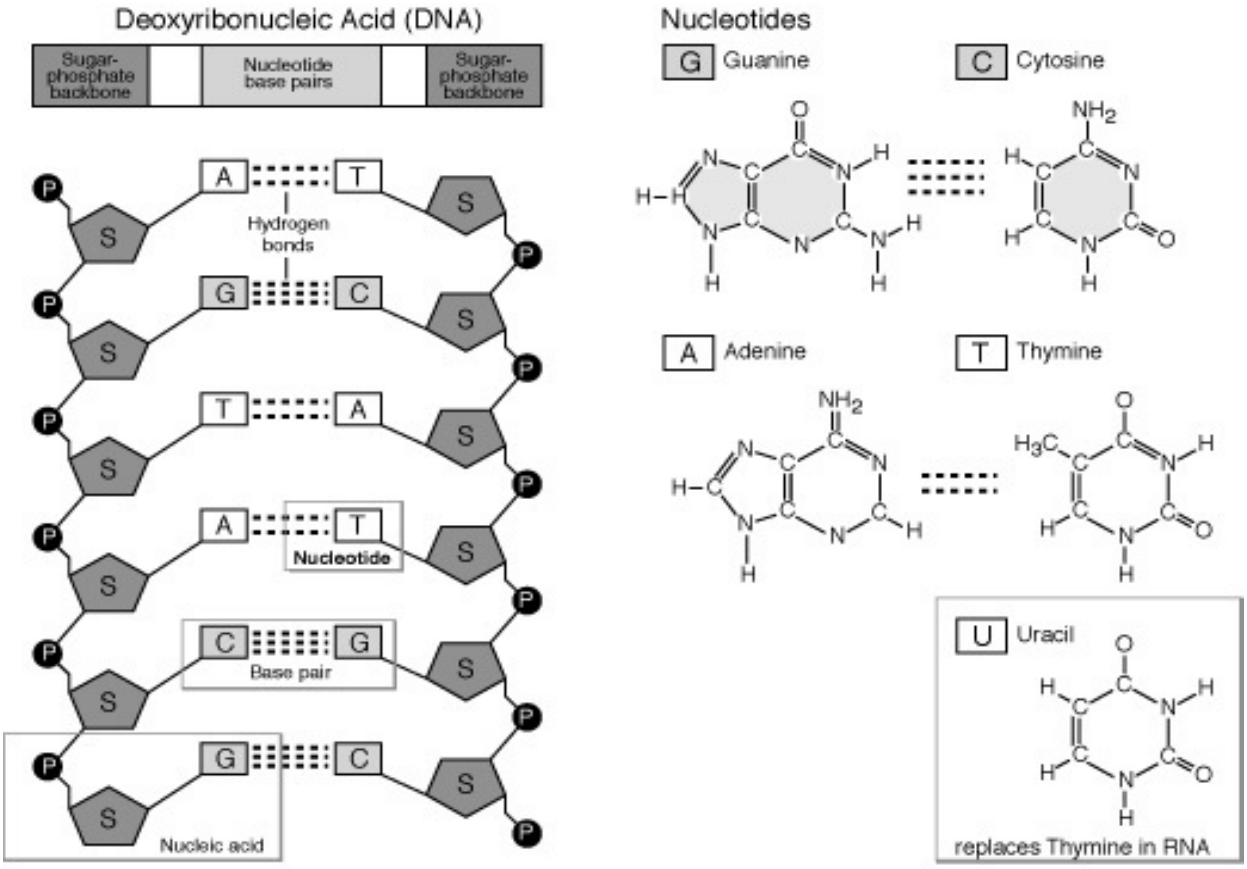
www.africanancestry.com – offers patri-clan and matri-clan testing (i. e. Y-DNA and mtDNA tests) and, with matches from their proprietary data bases, an identification of the African region from which the paternal or maternal ancestor probably came.

www.Ancestry.com – has recently announced that it will discontinue its Y-DNA and mt-DNA tests and database to concentrate on autosomal testing. I has a large data base of user-submitted family trees but had received criticism for not allowing users to work with the data behind their DNA matches.

www.AncestrybyDNA.com – this firm specializes in testing and interpreting a select group of SNPs scattered across many chromosomes and using the results to report on the mixture of ethnic origins in the individual's ancestry.

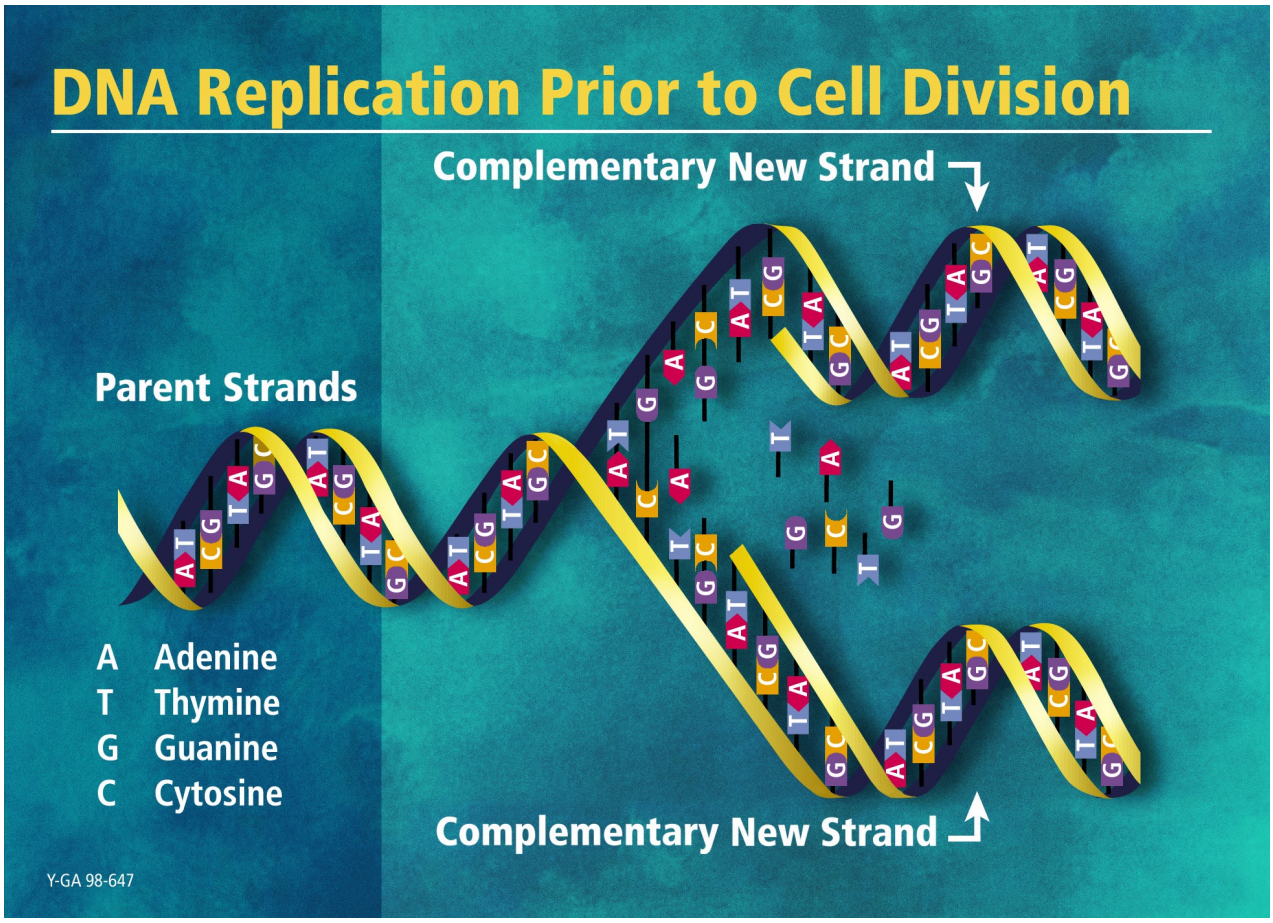
www.FamilyTreeDNA.com – this firm has very helpful customer support. The testing is done at the University of Arizona. It now hosts over 7800 surname and regional projects as well as serving individuals with a variety of DNA tests.

www.OxfordAncestors.com – offers several DNA tests but its speciality is the classification of European matrilineal descents into the Seven Daughters of Eve. Founded by Dr. Bryan Sykes.

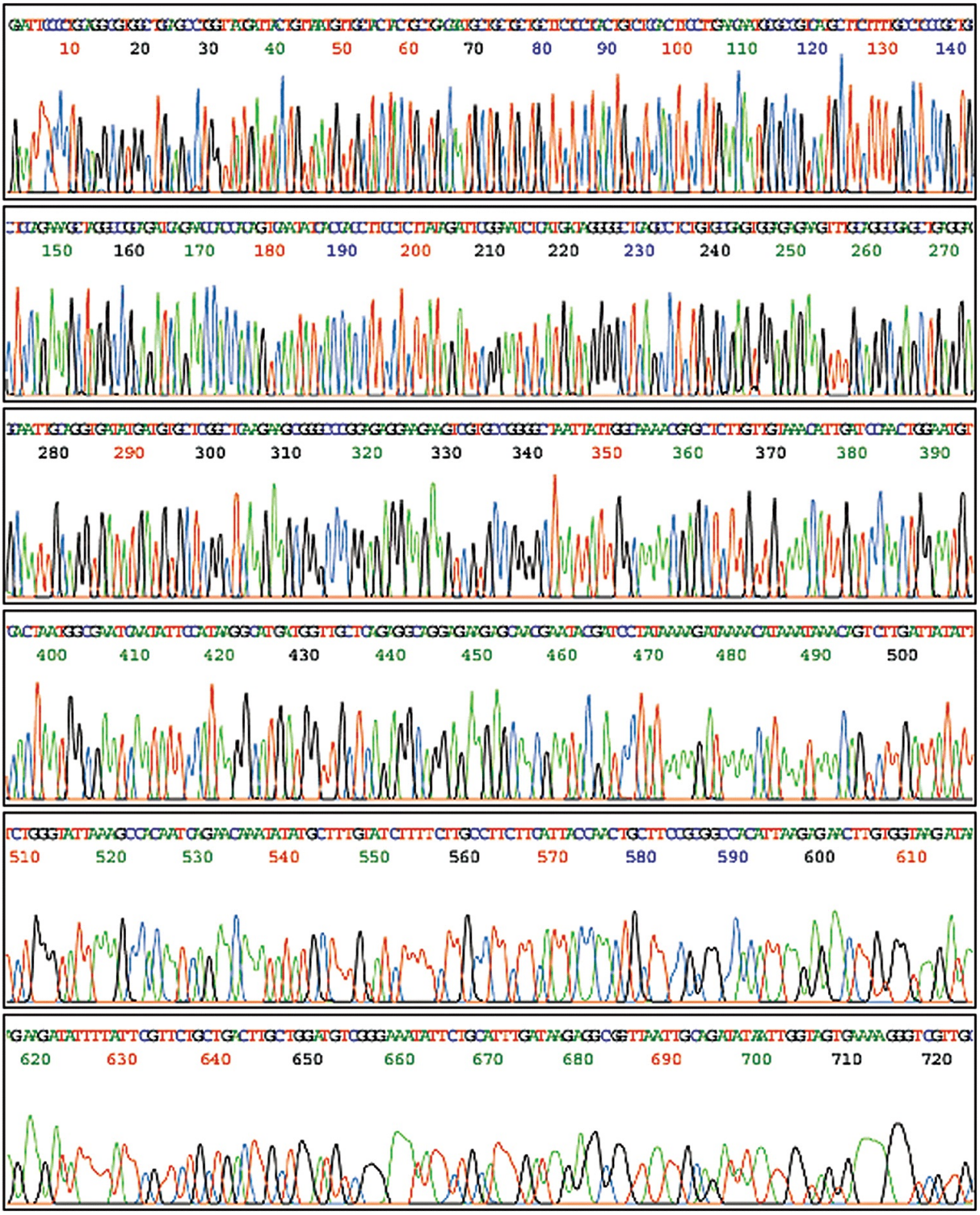



Nucleotide Diagram from National Human Genome Research Institute

DNA replication diagram from U.S. Dept. of Energy Human Genome Program
www.onrl.gov.hgmis



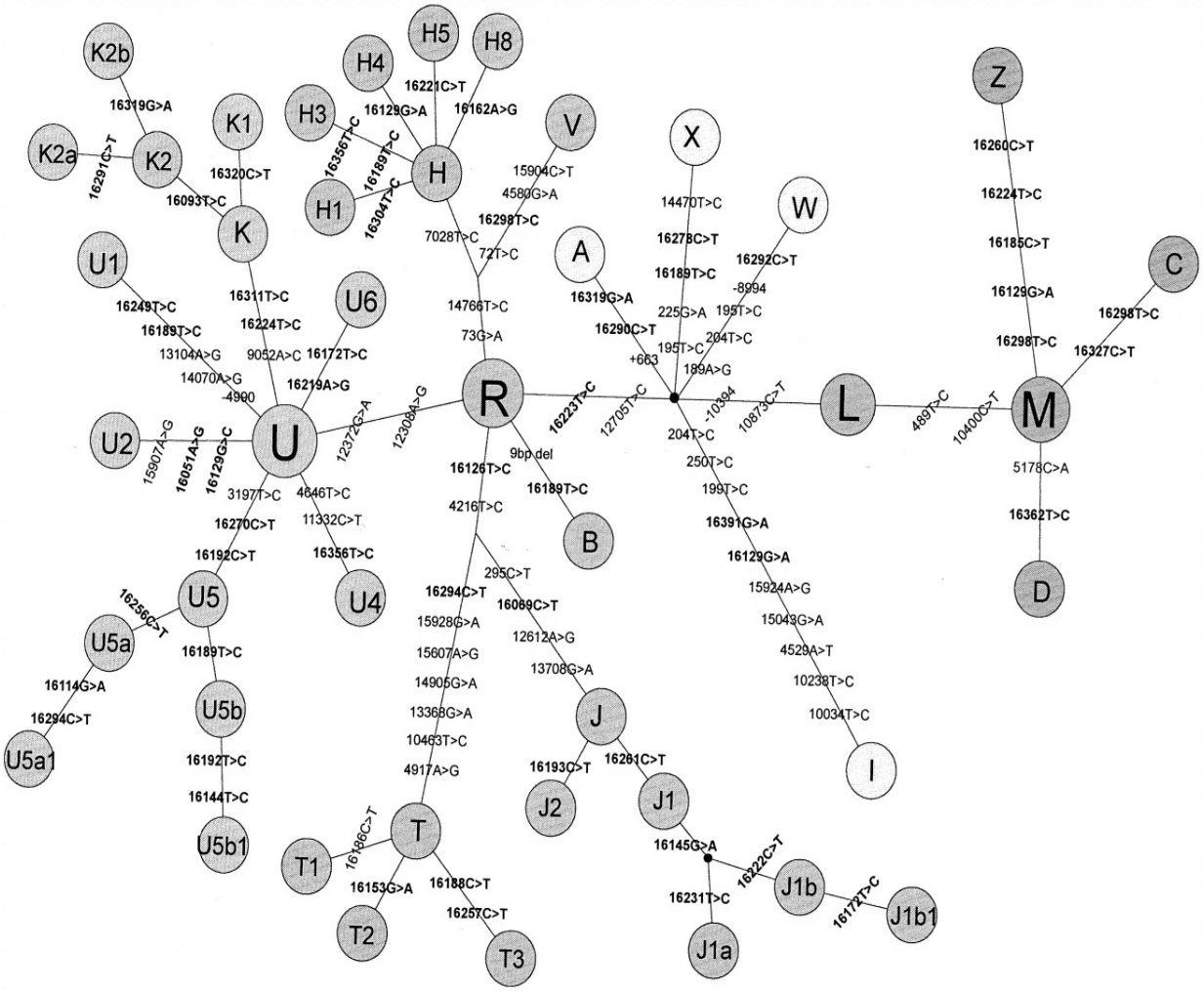
Genome Sequence Trace from U.S. Dept. of Energy Human Genome Program
www.ornl.gov/hgmis
This shows the interpreted result of a chromatographic scan of a DNA sample.



# One form of the mtDNA tree

(http://www.journals.uchicago.edu/AJHG/journal/issues/v68n3/002146/002146.fg2.html)

[This is now an old version. More current versions are less visually interesting, much more detailed and presented in outline form.]



Examples of possible DNA reports. Of course you may get a "suitable for framing" color certificate.

# Y-DNA Report for 37 locations

| Locus | DYS# | Alleles |
|---|---|---|
| 1 | 393 | 13 |
| 2 | 390 | 25 |
| 3 | 19 | 15 |
| 4 | 391 | 11 |
| 5 | 385 | 11-11 |
| 7 | 426 | 12 |
| 8 | 388 | 12 |
| 9 | 439 | 12 |
| 10 | 389-1 | 13 |
| 11 | 392 | 13 |
| 12 | 389-2 | 29 |
| 13 | 458 | 17 |
| 14 | 459 | 9-10 |
| 16 | 455 | 11 |
| 17 | 454 | 11 |
| 18 | 447 | 25 |
| 19 | 437 | 15 |
| 20 | 448 | 19 |
| 21 | 449 | 30 |
| 22 | 464 | 15-16-17-17 |
| 26 | 460 | 11 |
| 27 | Y-GATA-H4 | 11 |
| 28 | YCAII | 19-22 |
| 30 | 456 | 16 |
| 31 | 607 | 15 |
| 32 | 576 | 17 |
| 33 | 570 | 17 |
| 34 | CDY | 35-36 |
| 36 | 442 | 12 |
| 37 | 438 | 12 |

Your ntDNA results
Haplogroup - K2a10

Revised Cambridge Reference Sequence

HVR1 REFERENCE SEQUENCE

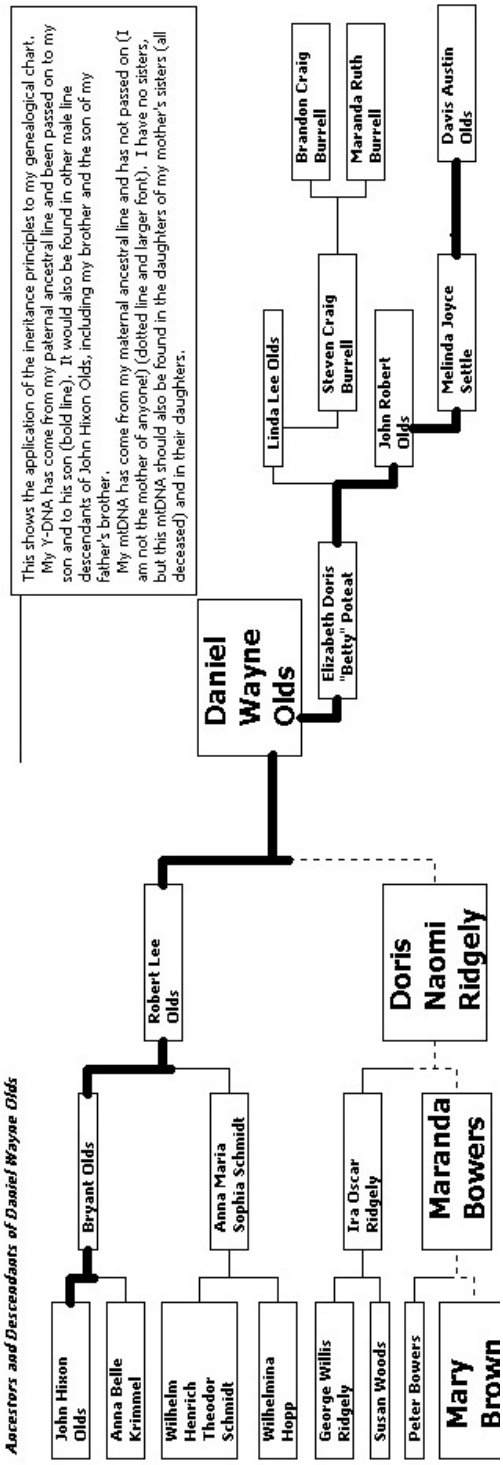
| Position | CRS | Your Result |
|---|---|---|
| 16224 | T | C |
| 16311 | T | C |
| 16519 | T | C |

HVR2 REFERENCE SEQUENCE

| Position | CRS | Your Result |
|---|---|---|
| 73 | A | G |
| 146 | T | C |
| 152 | T | C |
| 263 | A | G |
| 315.1 | C | (An insertion) |

CR REFERENCE SEQUENCE full mtDNA test
Show All Positions

| Position | CRS | Your Result | Position | CRS | Your Result |
|---|---|---|---|---|---|
| 709 | G | A | 12308 | A | G |
| 750 | A | G | 12372 | G | A |
| 1438 | A | G | 14167 | C | T |
| 1811 | A | G | 14766 | C | T |
| 2706 | A | G | 14798 | T | C |
| 3480 | A | G | 15326 | A | G |
| 4561 | T | C | | | |
| 4769 | A | G | | | |
| 7028 | C | T | | | |
| 8860 | A | G | | | |
| 9055 | G | A | | | |
| 9698 | T | C | | | |
| 9716 | T | C | | | |
| 10550 | A | G | | | |
| 11299 | T | C | | | |
| 11467 | A | G | | | |
| 11719 | G | A | | | |
| 11923 | A | G | | | |




## A Chart Showing the inheritance of Y-DNA and mtDNA

# A Typical Recent Ethnic Origins Report

| Ethnic Makeup | |
|---|---|
| European | 100% |
| European Coastal Islands | 38% |
| North Mediterranean Basin | 26% |
| European Northlands | 23% |
| Trans-Ural Peneplain | 8% |
| North Circumpolar | 5% |

## Interpreting Genetic Distance within Surname Project

**37 Markers**

Taken from a page copyrighted by Family Tree DNA

| **Distance** | **Relatedness** | **Explanation** |
|---|---|---|
| 0 | Very Tightly Related | A 37/37 match between two men who share a common surname (or variant) means they share a common male ancestor. Their relatedness is extremely close with the common ancestor predicted, 50% of the time, in 5 generations or less and over a 95% probability within 8 generations. Very few people achieve this close level of a match.<br><br>All confidence levels are well within the time frame that surnames were adopted in Western Europe. |
| 1 | Tightly Related | A 36/37 match between two men who share a common surname (or variant) indicates a close genealogical match. Very few people achieve this close level of a match, and it is within the range of most well-established surname lineages in Western Europe.<br><br>It's most likely that they matched 24/25 or 25/25 on a previous Y-DNA test, and the mismatch will be found within DYS576, DYS570, or CDY. |

| | | |
|---|---|---|
| 2<br>- | Related<br>- | A 35/37 match between two men who share a common surname (or variant) means they share a common male ancestor. The mismatch is likely within the range of most well-established surname lineages in Western Europe.<br><br>It is most likely that you matched exactly or closely on previous Y-DNA tests and the mismatch is within DYS439 or DYS385, DYS389i, 389ii, DYS458, DYS459, DYS449, DYS464, DYS576, DYS570, or CDY. |
| 3<br>- | Related<br>- | A 34/37 match between two men who share a common surname (or variant) means they share a common male ancestor. The relationship is likely within the range of most well-established surname lineages in Western Europe.<br><br>It is most likely that they matched exactly or closely on previous Y-DNA tests, and the mismatch is within DYS439 or DYS385, DYS389i, 389ii, DYS458, DYS459, DYS449, DYS464, DYS576, DYS570, or CDY. |
| 4<br>- | Probably Related<br>- | A 33/37 match between two men who share a common surname (or variant) means they may share a common male ancestor. This relationship should be confirmed with additional testing.<br><br>The only way to confirm the relationship is to test additional family lines and to find where the mutations took place. By testing additional family lines you can find the person in between. This 'in betweener' is essential for you to find. |
| 5<br>- | Only Possibly Related<br>- | A 32/37 match between two men who share a common surname (or variant) means that they may be related within the genealogical time frame, but additional evidence is needed to confirm the relationship.<br><br>If several or many generations have passed since the suspected common ancestor, it is possible that these two men are related. That would require that each line had experienced separate mutations and line would have experienced at least two mutations. The only way to confirm is to test additional family lines and find where the mutations took place. By testing additional family members you can find the person in between each of you. This 'in betweener' becomes essential for you to find, and without him the possibility of a match exists, but further evidence must be pursued. |

| | | |
|---|---|---|
| 6<br>- | Not Related<br>- | A 31/37 match between two men who share a common surname (or variant) means that they are not likely to be related within the genealogical time frame. The common surname is a coincidence.<br><br>If there is a strong family tradition of a relationship, it is distantly possible that these two men are related. That would require that each line had experienced separate mutations and line would have experienced at least two mutations. The only way to confirm the relationship is to test additional family lines and find where the mutation took place. By testing additional family members you can find the person in between the two men. This 'in betweener' becomes essential for you to find, and without him a genealogical relationship is unlikely. |
| Beyond 6<br>- | Not Related<br>- | A 31/37 match between two men who share a common surname (or variant) means that they are not likely to be related within the genealogical time frame. The common surname is a coincidence.<br><br>If there is a strong family tradition of a relationship, it is distantly possible that these two men are related. That would require that each line had experienced separate mutations and line would have experienced at least two mutations. The only way to confirm the relationship is to test additional family lines and find where the mutation took place. By testing additional family members you can find the person in between the two men. This 'in betweener' becomes essential for you to find, and without him a genealogical relationship is unlikely. |