

What you want or may need to know about DNA and Y-DNA

By Richard L. Tolman, Ph. D.¹

Since most of what is available to read about Y-DNA is largely unintelligible to someone familiar with the science or so watered down so as to be totally useless, I thought I would make my own attempt to explain the scientific background in this area.

DNA Replication

Few have a real understanding of how important DNA is—it is far more important than the family jewels—it defines everything about us and what we are and besides that carries with it lots of unused baggage in the way of heritage from our ancestors. The first thing to know is that it is almost unimaginably large (3.2 Billion base pairs). Consider what it might take in plans and construction to make all the pieces of our bodies (over 200 cell types) and then make all the machines (enzymes) that run the processes in our body as well as all the protective systems to keep everything running right. Most things in the body are made of protein (chains of aminoacids). To encode all the protein sequences that are required to make all these proteins only requires less than 5% of the total DNA in a cell. Think of DNA as a giant book; even if a cell only needs chapter 14 to run all its processes it still has the whole book in its nucleus. So in the DNA in our cells we have the important stuff (one twentieth of the total) plus a lot of other not-so-critical stuff that we have inherited from our ancestors or perhaps other stuff whose function the scientists have not figured out yet. This is how one of your kids can be the spitting image of great-uncle Fred—the kid has inherited some of the stuff we have carried along in our genome (DNA) but which we have not had the need to use in our lifetime. It also explains how my wife can be the grandniece of a fabulous musician and not have a lick of musical ability—life is not fair and no one said it would be.

Security—a very important concern for your DNA. These DNA sequences are made up of only 4 different building blocks (called A, C, G, T) and are fashioned into a double helix or duplex made up of two complimentary strands of DNA where all A's are paired with T's and G's are paired with C's. If one of these building blocks is damaged, changed by accident, or is somehow impaired it will not form a correct duplex with its partner in the duplex and will make a lump in the side of the double helix. Special repair enzymes roll down one of the grooves of the helix all the time and look for problems—if they find one, they fix it by excising the damaged building

¹ Richard L. Tolman, retired scientist/executive: Ph. D. in DNA bio-medicinal chemistry (University of Utah, 1969) with more than 30 years experience in DNA-related problems and research...and 15 years as a *pro bono* semi-professional genealogist/researcher.

block and putting in the right one. If they can't fix it for some reason, they arrange for a flag to be placed on the outside of the cell that says 'there is something wrong here—take out this cell' and one of your immune cells ('natural killer cells') comes along and kills the cell so that it can be replaced with a good one. So most problems with the DNA are fixed by the cell, but a select few are missed—these changes in sequence are called mutations. That's how important it is that the integrity and fidelity of the DNA is preserved. A change in a critical region (a 'lethal mutation') critically impairs the cell's function and will kill the cell; other mutations have minor effects on cell function that may be accommodated by the cell processes or the mutation may have no real effect on cell function. Cells which are defective in some way (have DNA mutations) produce defective progeny and therefore inevitably cause problems over time resulting in bad things like ageing and cancer. Despite all efforts, DNA undergoes some alteration over time.

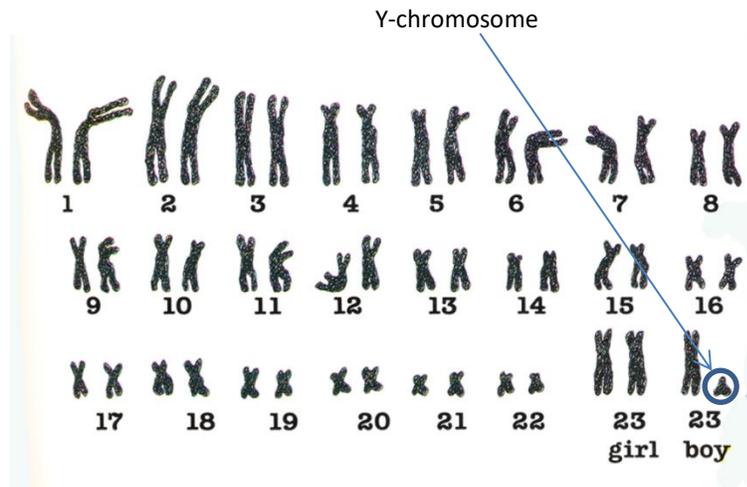


Figure 1.² Human chromosomes (1 to 22 arranged from largest to smallest; chromosome 23 here is actually #3)

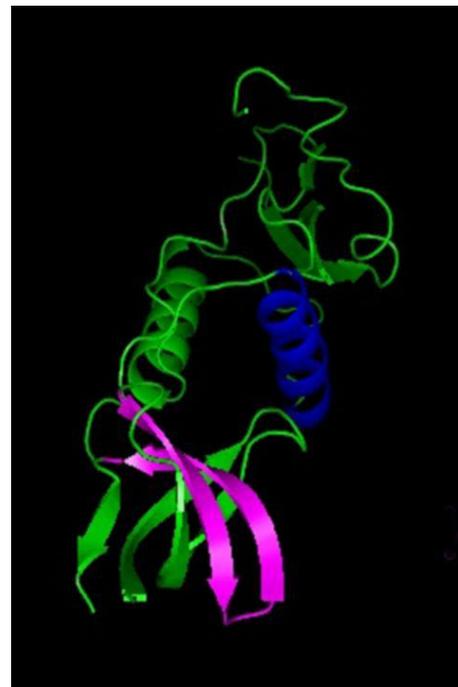


Figure 2.³ A folded protein

To protect the DNA even more effectively, your body has other techniques. One is supercoiling. Consider a balsa-wood model airplane with a rubber band motor. At rest the

² Original from Balkwill, Fran (illust. By Mic Rolph) *Amazing Schemes within your Genes* (Carolrhoda Books: London, England, 1993), p. 15 (with modification).

³ Online at Wikipedia, search: 'protein folding'.

rubber band is flat and straight. As you wind up the propeller, the rubber band knots up and then the knots knot up, until in the extreme case you have a big round knot of solid rubber in the middle of the model airplane. This is called supercoiling and this is what your DNA does to help protect itself from foreign intervention from ultraviolet rays or other similarly nasty things that could damage (mutate) the sequences. Figure 1 shows an electromicrograph of each of the human chromosomes, including the Y-chromosome (#3). You can see that it looks like a big Y-shaped knot. To read the DNA the cell has a big complex of machines that work together to unwind, read, and make a transcript of the sequence. The very clever unwinding enzymes are called gyrases (bacterial) or topoisomerases.

Think about how complicated the DNA must be in order to tell the cell how to make a given protein or enzyme. If you read instructions for making a model airplane on a piece of paper (2-dimensional)—it contains pieces to be cut out and also instructions about how to build the airplane (which will ultimately be 3-dimensional). This is what the DNA does as well—in a long chain of characters (2-dimensional) it has the code for the aminoacids that have to be linked together to make the protein and also interspersed in the chain, instructions on how to make those aminoacid chains into a 3-dimensional object that will work. Figure 2 shows a typical protein. Putting this baby together would be very complicated and a lot harder than building a model airplane. Scientists today only understand a few of the ways DNA can encode these complicated instructions.

DNA Sequencing

There are three kinds of commercial DNA sequencing available. Autosomal DNA (**atDNA**), Mitochondrial DNA (**mtDNA**), and **Y-DNA** (handled separately).

atDNA is the most common type of DNA-sequencing available (sold by Ancestry and others). It is useful for finding cousins and biological parents; it is not gender-specific. Every son/dau inherits one-half of the genetic material from each parent; hence your gdau will have 800 Million base pairs (1/4) of your DNA randomly distributed along her genome. You will possess only 6.25% of your great grandparents DNA. Therefore **atDNA** has only limited genealogical utility since it can only go back 4-5 generations.⁴

mtDNA sequencing can be determined for men and women, but it only follows the matriarchal line; you inherit it only from your mother. In your individual cells there is a nucleus in the center where the DNA resides. Floating about in the rest of the cellular space (cytoplasm) are mitochondria—important little organelles which respire (use oxygen) and

⁴ Bettinger, Blaine *Guide to DNA Testing and Genetic Genealogy* (Cincinnati, OH: Family Tree Books, 2015); probably the best explanation and visuals available about atDNA.

generate the energy which runs the cell. Since they are so important they have their own tiny pieces of DNA (**mtDNA**; 17,000 base-pairs) that they inherit exclusively from Mom. These are genealogically useful for tracing the matriarchal line (so-called *umbilical line*; your mother's mother's mother, etc.). Considering how small the DNA is, one needs a nearly perfect match to prove a matriarchal lineage. To find out more about mitochondrial genetics, I recommend reading the immensely readable and enjoyable book by Bryan Sykes *The Seven Daughters of Eve*.⁵ Not to minimize its utility, but you need to know a lot about your ancestry to use **mtDNA**; since in following the matriarchal line, each generation has a surname change.

Y-DNA

And the third type of DNA sequencing is **Y-DNA**, which has tremendous genealogical utility. We have 23 pairs of chromosomes (pairs because one was originally from mom (#3 = XX) and the other from dad (#3 = XY)). The Y-chromosome is the smallest of all the chromosomes (about 59 million base-pairs; information-carrying building blocks); it encodes all the masculine receptors and carrier proteins and other male enzymes necessary to become a biological male. The sequence of Y is obtained entirely from dad; mom cannot help because she does not have these DNA sequences (she has two X-chromosomes). All babies *in utero* develop female genitalia until the Y-chromosome is read and the Y-proteins are made which then masculinizes the fetus (Wolffian differentiation).

The base-pair sequence of **Y-DNA** can be traced back *for thousands of years*. Roughly 2 mutations occur each generation and occasionally one of these mutations occurs in one of the key monitoring loci/marker sequences (when you have your **Y-DNA** sequenced they sequence the whole thing, but monitor specific highly-variable loci—typically 37 sites for the cheapest test—it can be expanded to 67 sites or 111 sites for more money). Based on the mutations that have occurred at each of these loci, your **Y-DNA** is placed in a 'Haplogroup' (=category) by those determining the base sequence. Today Y chromosome sequencing is done only by FamilyTreeDNA.com and surname tables are maintained online⁶ to collect together by surname all Y-DNA sequences and loci information (STR's) in a panel of 37, 67 or 111 loci/markers. Individuals with the same surname that are genealogically related will have a related panel of STR values (see below) and will be in the same Haplogroup.

The **Y-DNA** sequence includes the coding regions (information about the critical protein sequences for the masculine proteins) and also instructions for assembling them. The coding

⁵ Sykes, Bryan *The Seven Daughters of Eve* (New York, NY: Norton, W. W. & Co., Inc, 2001).

⁶ Surname DNA Project—Y-DNA Colorized Chartonline at

<https://familytreedna.com/public/surname?iframe=colorized>; these charts are no longer public and are available only to those who have uploaded data and can sign in to the site (accessed 7 Jul 2018).

regions for the aminoacids cannot tolerate many mutations as this would affect the working of the enzymes/proteins. The in-between regions which can encode instructions are much more tolerant to change. They include *short tandem repeats* (STR's), repeated nonsense sequences (spacers) that are important to genealogists. **The STR value at each of the loci/markers in the Y-DNA panel is the number of times the nonsense sequence is repeated.** So you can see how diagnostic it may be to examine this huge sequence and count the number of times the nonsense sequence is repeated for each one of these loci/markers—these numbers are handed down religiously from father to son because the son's Y-chromosome was copied from his father's. In order to decide if two individuals with the same surname share a common ancestor, they must have a small *genetic distance* or divergence in DNA sequences, that is—there must be a small number of differences between the STR values in the panel of loci/markers. A genetic distance of 0 would mean that the two DNA sequences are identical.

I hope this sheds some light on this terrifically interesting topic.

DNA Privacy Issues

It is possible to get even deeper into the morass of DNA research and genealogy. You can map your chromosomes to find out exactly which pieces of DNA you share with relatives/ancestors. I will not go into this any more than to say it may be important?/interesting to know how many centimorgans (a measure of DNA distance in contiguous genes) and on which chromosome you share DNA with others. You do this 'chromosome mapping' by uploading your genome (DNA sequence) to a website called GEDmatch.com, but you are not sharing it with the world. Some people are concerned that once you release your DNA sequence into the 'public' somewhere, your privacy may be invaded. That is—a prospective employer could access your genome and determine that you have genes associated to autism,⁷ alcoholism,⁸ violence,⁹ Alzheimer's¹⁰ (take your choice) and decide not to hire you.

It is easy to become paranoid about things we don't understand well. First of all, we are ALL going to have genes that may be linked to one or more of those bad traits/diseases. Second of all, it is or soon will be against the law to do these kinds of studies for malicious reasons. Third of all, we already know from the great volume of cancer genetics that has been

⁷ Chaste, Pauline and Marion Leboyer 'Autism risk factors: genes, environment, and gene-environment interactions' *Dialogues Clin Neurosci* (Sep 14, 2012) 281-292.

⁸⁸ 'Genetics of Alcoholism' online at www.addictioncenter.com/alcohol/ (accessed 7 August 2018).

⁹ Garcia-Arocena, Dolores 'The Genetics of Violent Behavior' online at www.jax.org/news-and-insights/jax-blog/2015/december (accessed 7 August 2018).

¹⁰ Mayo Clinic Staff 'Alzheimer's genes: Are you at risk?' online at www.mayoclinic.org/diseases-conditions/alzheimers-genes (accessed 7 August 2018).

done that nature makes nothing simple. We used to believe that there would be a ‘liver-cancer-gene’, but this has turned out to be hopelessly naïve. Women should be very concerned if they have the BRCA2 gene mutation since this has been linked to the development of breast cancer. But it doesn’t mean that you will get breast cancer if you have the BRCA2 mutation, it only means that you have an *increased probability* of developing breast cancer.¹¹ There are a host of genes linked to the development of Alzheimers—if you have a couple of them it may increase your probability of developing the disorder. But it is far too complex (and unfair) to make judgements about people based on genes they may possess. There are already over 60 genes that are linked to specific genetic diseases and it will soon be possible to warn couples if they have conceived a child that may with high probability inherit one of these genetic diseases.¹² Fourth of all, exactly for the same reason we should share our Family Tree data (dead people only) with the world—only by sharing can we learn more. We need to share DNA data with hospital researchers and other genetic researchers so they have a chance to figure out our vulnerabilities and how to deal with them and more about what makes us what we are. Come on in the water’s fine.

¹¹ Breast Cancer Genetics, online at <https://www.breastcancer.org/risk/factors/genetics> (accessed 7 Aug 2018).

¹² Lipkin, Steven Monroe The Age of Genomes: Tales from the Front Lines of Genetic Medicine (Boston, MA: Beacon Press, 2016); a terrific, eye-opening book.