Y-DNA Tutorial

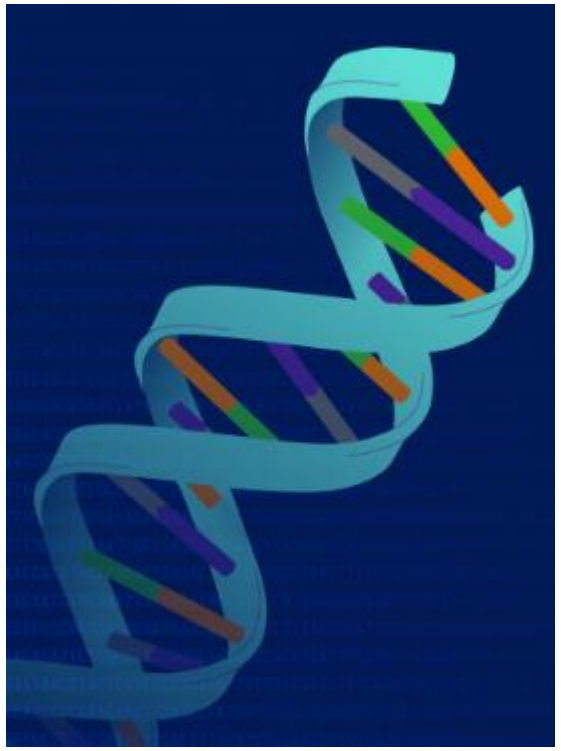
Herriott Surname Project

September 2020

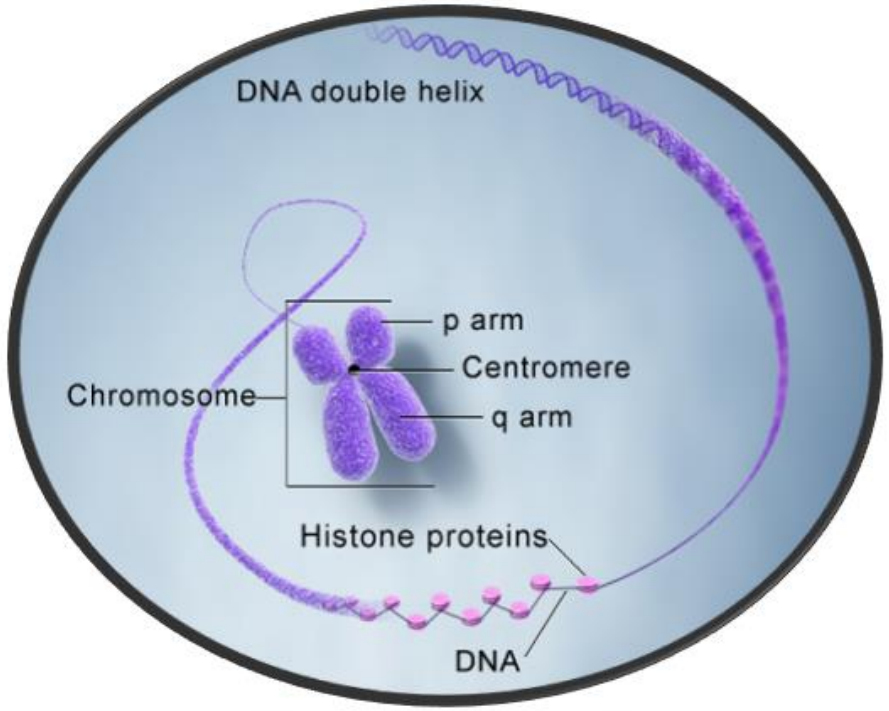
What is DNA made of?

DNA – Deoxyribonucleic Acid

- DNA is made of chemical building blocks called nucleotides. These building blocks are made of three parts: a phosphate group, a sugar group and one of four types of nitrogen bases. To form a strand of DNA, nucleotides are linked into chains, with the phosphate and sugar groups alternating.
- The four types of nitrogen bases found in nucleotides are: adenine (A), thymine (T), guanine (G) and cytosine (C). The order, or sequence, of these bases determines what biological instructions are contained in a strand of DNA. For example, the sequence ATCGTT might instruct for blue eyes, while ATCGCT might instruct for brown.
- The complete DNA instruction book, or genome, for a human contains about 3 billion bases and about 20,000 genes on 23 pairs of chromosomes.



Chromosomes



Chromosomes are paired threadlike "packages" of long segments of DNA contained within the nucleus of each cell. In humans there are 23 pairs of chromosomes. In 22 pairs, both members are essentially identical, one deriving from the individual's mother, the other from the father. The 23rd pair is different. In females this pair has two like chromosomes called "X". In males it comprises one "X" and one "Y," two very dissimilar chromosomes. It is these chromosome differences which determine sex.

The Y-Chromosome

Human sex is determined by the X and Y chromosomes. A female has 2 X-Chromosomes and a male has an X and a Y-Chromosome. When a child is conceived it gets one chromosome from its mother and one chromosome from its father. The chromosome from the mother will always be an X, but the chromosome from the father may be either X or Y. If the child gets the X she will a girl,if the child gets the Y he will be a boy.

The Y-Chromosome

This Y-Chromosome has certain unique features:

- The presence of a Y-Chromosome causes maleness. This little chromosome, about 2% of a father's genetic contribution to his sons, programs the early embryo to develop as a male.
- It is transmitted from fathers only to their sons.
- Most of the Y-Chromosome is inherited as an integral unit passed without alteration from father to sons, and to their sons, and so on, unaffected by exchange or any other influence of the X-Chromosome that came from the mother. It is the only nuclear chromosome that escapes the continual reshuffling of parental genes during the process of sex cell production.

It is these unique features that make the Y-Chromosome useful to genealogists.

Testing the Y-Chromosome

The Y-Chromosome has definable segments of DNA with known genetic characteristics. These segments are known as **Markers**. These markers occur at an identifiable physical location on a chromosome known as a **Locus**. Each marker is designated by a number (known as **DYS#**), according to international conventions. You will often find the terms Marker and Locus used interchangeably, but technically the Marker is what is tested and the Locus is where the marker is located on the chromosome.

Testing the Y-Chromosome

Although there are several types of markers used in DNA studies, the Y-Chromosome test uses only one type. The marker used is called a **Short Tandem Repeat (STR).** STRs are short sequences of DNA, (usually 2, 3, 4, or 5 base pairs long), that are repeated numerous times in a head-tail manner. The 16 base pair sequence of "gatagatagatagata" would represent 4 repeats of the sequence "gata". These repeats are referred to as **Allele**. The variation of the number of repeats of each marker enables discrimination between individuals.

Reading the Test Results

The table to the right is a shortened version of the actual table used to show our DNA test results. It shows 12 of the 25 markers that most of the participants had tested.

Marker	1	2	3	4	5	6	7	8	9	10	11	12	
	DYS#												
Part ID#	393	390	19*	391	385a	385b	426	388	439	389i	392	389ii	Ancestor #
3947	13	26	14	11	12	14	12	12	11	13	13	29	0001

The numbers (1-12) across the top of the table are the marker numbers. They have no significance other than as an easy way to refer to the marker. Note: FamilyTree DNA refers to these numbers as Locus. The second set of numbers across the top of the matrix are DYS# (the actual marker names).

The numbers down the left side of the table identify the participant in the DNA project. The numbers down the right side of the table identify the participant's oldest known ancestor. The rest of the numbers are the Allele (the number repeats) for each participant at the specified marker.

What Does it Mean

An individual's test results have little meaning on their own. You cannot take these numbers, plug them into some formula and find out who your ancestors are. The value of the test results depends on how your results compare to other test results. And even when you match someone else, it will only indicate that you and the person you match share a common ancestor.

Depending on the number of markers tested and the number of matches it will indicate with a certain degree of probability how long ago this common ancestor existed. It will not show exactly who this ancestor is.

What Does it Mean

As discussed above, the Y-Chromosome is passed from father to son. The vast majority of the time the father passes an exact copy of his Y-Chromosome to his son. This means that the markers of the son are identical to those of his father. However on rare occasion there is a mutation or change in one of the markers. The change is either an insertion or a deletion. An insertion is when an additional repeat is added to a marker. A deletion is when one of the repeats is deleted.

Mutations

Mutations occur at random. This means it is possible for two distant cousins to match exactly on all markers while two brothers might not match exactly. Because of the random nature of mutations we must use statistics and probability to estimate the **Time to the Most Recent Common Ancestor (TMRCA).** The actual calculations of TMRCA are mathematically complex and depend on knowing the rate of mutation and the true number of mutations. At this time there is not enough data to accurately determine either of these factors so certain assumptions have to be made.

TMRCA Calculations

The simplest and one of the most commonly used models makes the following assumptions:

- **Rate of Mutation = .002.** This assumes that any given marker has a .002 chance of mutating with each generation. In other words, we could expect any marker to mutate once in 500 generations. The rate of .002 is considered conservative and is the average of a number of studies. It will result in a TMRCA that is longer than higher mutation rates.

- **Number of mutations:** This model counts any change in a marker as a single mutation. Each marker is scored as either a match or a non-match. If a marker does not match it is assumed to be a single mutation. This method a counting mutations may result in underestimating the TMRCA.

Cumulative Probability Table

Based on the above assumptions we derive the cumulative probability table. This table simply list the number of generations corresponding to the 50%, 90% and 95% probability levels for various numbers of matches.

Match		50%	90%	95%	95% Confidence Interval
12-0	Match exactly at all 12 markers	14	48	62	1-77
11-1	11 exact matches, 1 mismatch	37	85	103	5-121
10-2	10 exact matches, 2 mismatch	61	122	144	14-165
25-0	Match exactly at all 25 markers	7	23	30	0-37
24-1	24 exact matches, 1 mismatch	17	40	48	2-57
23-2	23 exact matches, 2 mismatch	28	56	66	6-75

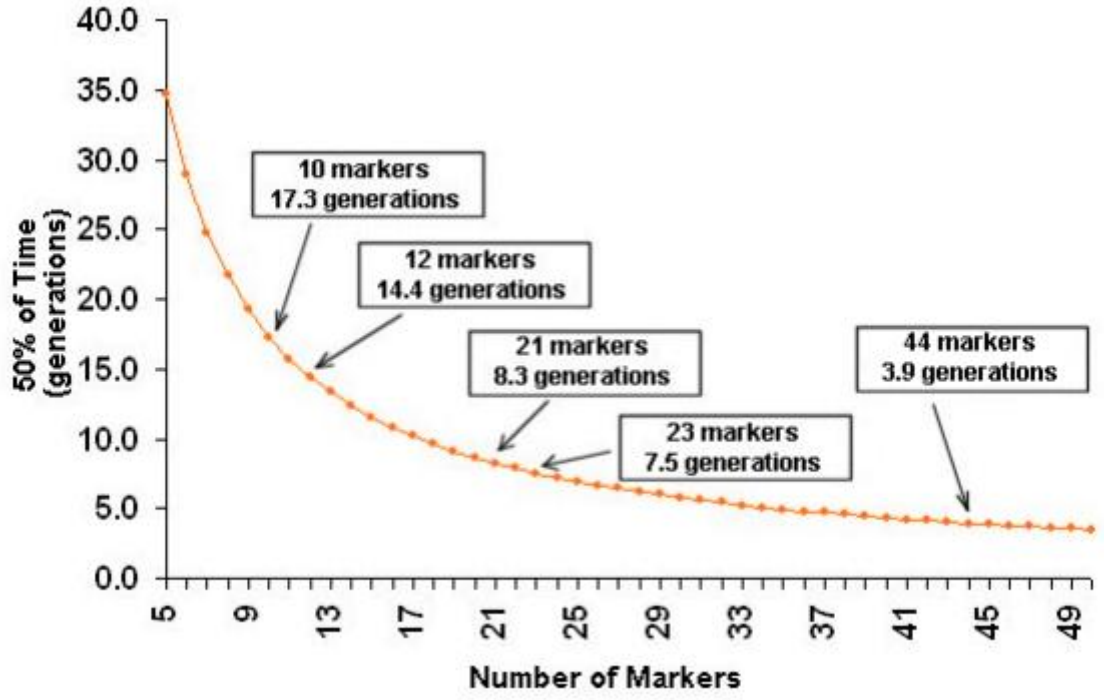
The TMRCA for 12 markers assumes that there are ONLY 12 markers available for testing. If there are only 12 markers and you match 12 for 12, there is a 50% probabilty that you share a common ancestor within 14 generations

The TMRCA for 25 markers assumes that there are ONLY 25 markers available for testing. If there are only 25 markers and you match 25 for 25, there is a 50% probabilty that you share a common ancestor within 7 generations

This table tells us that if we match on 24 of 25 markers there is a 50% probability that the most recent common ancestor is 17 generations or less, a 90% probability that TMRCA is 40 generations or less, and a 95% probability that TMRCA is 48 generations or less. The 95% Confidence Interval is the upper and lower range of values that encompass 95% of the probability for the TMRCA. If we match on 24 of 25 markers, 95% of the possible TMRCA values fall
between 2 and 57 generations.

Cumulative Probability Table

As you can see from the above table more markers reduce the number of generations to TMRCA. The Chart below shows how increasing the number of markers tested, decreases the number of generation to TMRCA when all markers match.



Putting It All Together

DNA testing can be a valuable tool in genealogical research when it is combined with conventional research. Test results can be used to confirm a suspected connection between two families or disprove a connection. Although it is impossible to pinpoint a common ancestor from the test results alone, with a proper paper trail you may be able to do so.

Useful Definitions

Allele: One of the variant forms of a gene at a particular locus, or location, on a chromosome. Different alleles produce variation in inherited characteristics. For STR markers, each allele is the number of repeats of the short base sequence.

Base Pair: Two bases that form a "rung of the DNA ladder." A DNA nucleotide is made of a molecule of sugar, a molecule of phosphoric acid, and a molecule called a base. The bases are the "letters" that spell out the genetic code. In DNA, the code letters are A, T, G, and C, which stand for the chemicals adenine, thymine, guanine, and cytosine, respectively. In base pairing, adenine always pairs with thymine, and guanine always pairs with cytosine.

Useful Definitions

Chromosome: One of the threadlike "packages" of genes and other DNA in the nucleus of a cell.

DNA: The chemical inside the nucleus of a cell that carries the genetic instructions for making living organisms.

DYS#: D=DNA, Y=Y chromosome, S=a unique DNA segment. A label for genetic markers on the Y chromosome. Each marker is designated by a number, according to international conventions. At present, virtually all the DYS designations are given to STR markers (a class often used in genetic genealogy).

Useful Definitions

Gene: The functional and physical unit of heredity passed from parent to offspring. Genes are pieces of DNA, and most genes contain the information for making a specific protein.

Genome: All the DNA contained in an organism or a cell, which includes both the chromosomes within the nucleus and the DNA in mitochondria.

Locus: A point in the genome, identified by a marker, which can be mapped by some means. It does not necessarily correspond to a gene. A single gene may have several loci within it (each defined by different markers) and these markers may be separated in genetic or physical mapping experiments. In such cases, it is useful to define these different loci, but normally the gene name should be used to designate the gene itself, as this usually will convey the most information.

Useful Definitions

Marker: Also known as a genetic marker, a segment of DNA with an identifiable physical location on a chromosome whose inheritance can be followed. A marker can be a gene, or it can be some section of DNA with no known function. Because DNA segments that lie near each other on a chromosome tend to be inherited together, markers are often used as indirect ways of tracking the inheritance pattern of genes that have not yet been identified, but whose approximate locations are known.

Microsatellite: Repetitive stretches of short sequences of DNA used as genetic markers to track inheritance in families.

Mutation: A permanent structural alteration in DNA.

Useful Definitions

Short Tandem Repeats (STR): A genetic marker consisting of multiple copies of an identical DNA sequence arranged in direct succession in a particular region of a chromosome. Occasionally, one will mutate by the gain or loss of one repeat. (Also known as microsatellite)