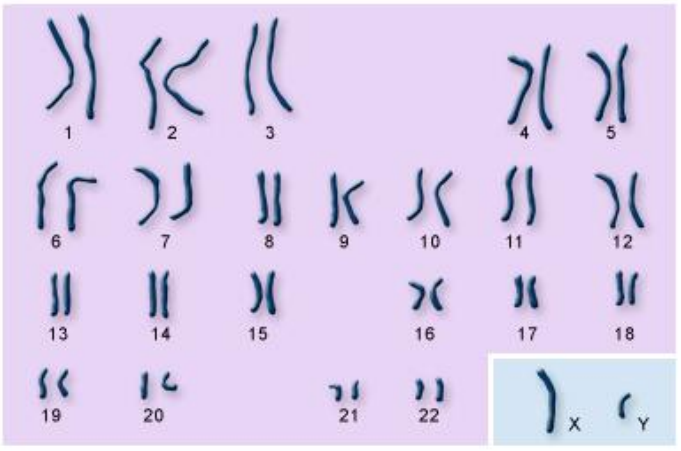
Advanced Y-DNA

Lesson 1. Y-DNA in Detail

Objective: Understand advanced Y-DNA topics and how it affects your genealogy
Tools: This lesson will allow you to delve deeply into your Y-DNA details and know how it affects your paternal genealogy.

- Words and phrases in **bold** indicate important terminology. Please consult the glossary included with the course material.
- Click any entry in the Table of Contents to jump to that part of the lesson.
- Class notes are at the end of the lesson.



U.S. National Library of Medicine

Lesson written by Gale French
Version 1.0 (Apr. 2021)
Layout and Design Gale French

Contents

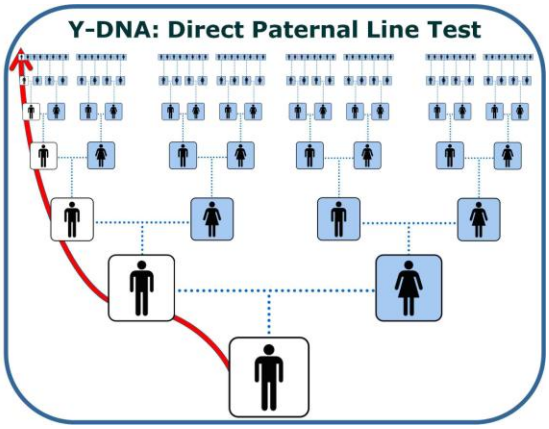
Section 1. Y-DNA Basics Review

This section is a review of Y-DNA (taken from the Basic Y-DNA course at DNAAdoption). It sets a foundation for the advanced information for previous students or those who didn't take the Basic course.

Types of DNA

There are four basic types of DNA used in genealogical testing:

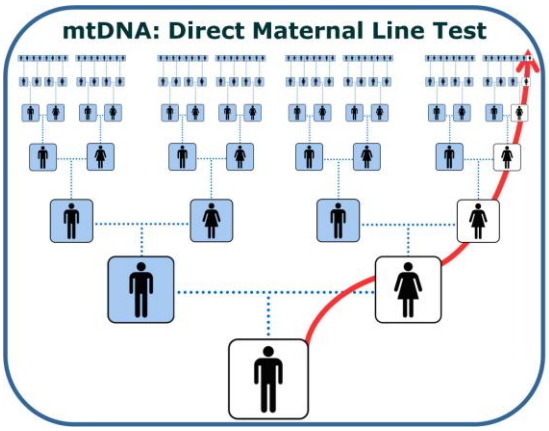
1. Y-DNA - DNA derived from the Y Chromosome



The Y-chromosome is only passed down from a father to a son, unchanged from generation to generation (unless a random naturally occurring change called a **mutation** happens). The result of that mutation (called a **marker**) can be traced through many generations as that man will pass it on to his sons, and to every male in the family for many generations. On average, mutations in any marker are estimated to occur only once in every 500 generations.

The Y-DNA **STR** test offers a clear path from you to a known, or likely, direct paternal ancestor(s).

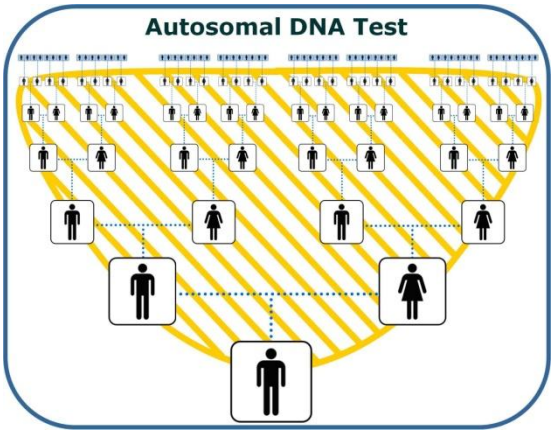
2. mtDNA - **Mitochondrial** DNA



mtDNA is the line that follows your mother's maternal ancestry. This line consists entirely of women, although both men and women have their mother's mtDNA. This means that fathers do not pass on their mtDNA to their children.

The mtDNA test can trace your mother, her mother, her mother's mother, and so forth, and offers a clear path from you to a known, or likely, direct maternal ancestor.

3. atDNA - **Autosomal** DNA



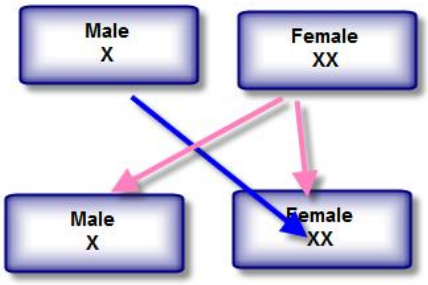
atDNA is inherited randomly; you will inherit segments that your mother passed on to you, some that your father passed to you, some from your grandparents and maybe even some from your paternal third great-grandmother.

This test is designed to find relatives on any of your ancestral lines within the last 5 generations.

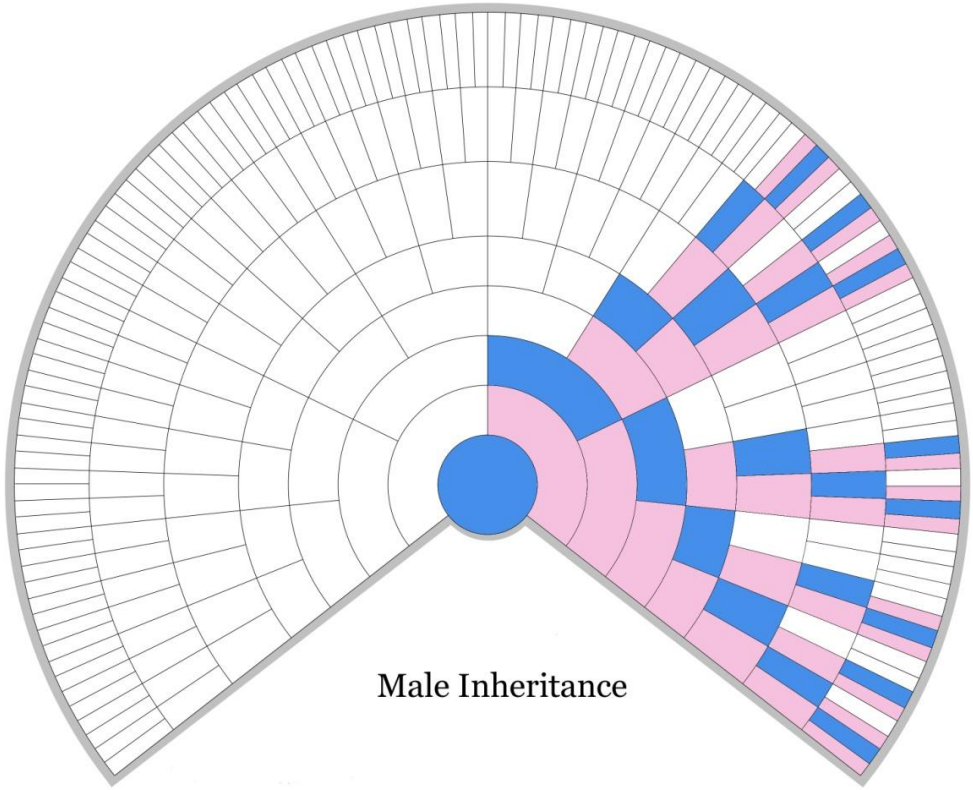
4. X-DNA - **X-Chromosome** DNA

What is X-DNA?

X is not a separate test; it is reported as part (Chromosome 23) of the Autosomal test. It is really quite simple. A man passes down his X chromosome only to his daughter. A woman passes down her X chromosome to both her sons and daughters. Now let's reverse the picture. A man gets his chromosome from his mother, not his father. A woman gets hers from both her mother and father. We are saying the same thing, but one description might make more sense to you than the other.



A male never gets an X from another male. The chart below[1] shows the X-Chromosome inheritance pattern (where blue is male and pink is female).



This lesson will concentrate on the **Y-DNA**.

[1] Source: Blaine Bettinger

STR Markers

A chromosome contains sequences of repeating nucleotides known as **Short Tandem Repeats** (**STRs**). The number of repetitions varies from one person to another and a particular number of repetitions is known as an **allele** of the marker. An STR on the Y-chromosome is designated by a **DYS** number (**D**NA **Y**-chromosome **S**egment number). For an STR test, short segments of DNA are measured. The number of repeats in that short sequence, represent a marker value for a particular DYS. The outcome of the Y-DNA test is presented as a series of marker values and the results, or alleles, as in the following 12-marker example:

DYS#	393	390	19	391	385a	385b	426	388	439	389-1	392	389-2
Alleles	13	24	14	10	11	14	12	12	12	13	13	29

Someone with a result of 12 at DYS426 would have the following sequence:

…TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGAC…

At each location on the Y-chromosome, there is the potential for some variation; at DYS426, the variation consists of 7 to 18 repeats of the DNA sequence "GTT", with 12 being the most common. [2]

The more markers that are tested, provide a better possibility of identifying paternal matches for more generations. The fewer mismatches testers show, the shorter the time that has elapsed since their **Most Recent Common Ancestor** (**MRCA**). The goal is to find a person(s) that match the most markers with you. That means a very close relationship. It is recommended that you start with a minimum of 37 markers.

Y-DNA Genetic Distance

When talking about two or more Y-Chromosome STR haplotypes, **genetic distance** is the total number of differences, or mutations, between two sets of results. In general, it is found by summing the differences between each STR marker. For example, kit B111 and B222 have allele values of 29 and 28 respectively at DYS389-2. This is a difference of 1 {29-28= 1}. Because this is the only difference in their Y-DNA12 profiles, their genetic distance is 1.

Kit	Surname	DYS393	DYS390	DYS19	DYS391	DYS385	DYS426	DYS388	DYS439	DYS389-1	DYS392	DYS389-2
B111	Smith	12	23	14	10	17-18	11	12	11	12	14	**29**
B222	Smith	12	23	14	10	17-18	11	12	11	12	14	**28**
B333	Smythe	12	22	14	10	16-17	11	12	11	12	14	29

Pop Quiz #1

What is the Genetic Distance between kit B222 and B333? (*be careful*!!)

[2] Source: FT-DNA

There are a couple of "odd" cases to consider for computing Genetic Distance:

1. "No Calls" - occurs when a SNP being analyzed has insufficient data to be confidently given a genotype value
2. Multi-Value Markers

No Calls

If a marker has a "No Call" (a value of 0), that marker is counted as **1** difference (e.g., **NOT** 14-0=14) when compared to other markers with >0 values.

Multi-Value Markers

For multi-value markers (Palindromic markers will be covered in "Palindromes" on page 15), FTDNA counts a difference for two types of changes: a Mismatch and a Copy number change. **Mismatches** occur when the compared markers do not match. For example, person A might have DYS464 = 14-15-15-16 and person B might have DYS464 = 14-15-16-18. This counts as a single difference in the two men's Genetic Distance. **Copy number changes** are when one male has more copies of the STR marker than the other male. For example, one man might have DYS464 = 14-15-16-17 and the other might have 14-15-15-16-16-17. This also counts as a single difference toward the two men's total Genetic Distance. When two men are compared and they show both a Mismatch and a Copy number change, it is counted as a genetic distance of 2. For example, one might have DYS464 = 14-14-15-16 and the second person DYS464 = 14-15-15-16-16-17.

MRCA

This table shows the probability that your common ancestor (MRCA) lived about this number of generations ago. The percent chance of a common ancestor within a given time is not an exact science, thus the calculations vary as the genetic distance grows. The table shows the probability for multiple markers (the key for the 2nd row is: #markers/genetic dist. (e.g., **25/1** means 25 markers test with one difference):

		Percent Chance of a Common Ancestor in Genealogical Terms, Based on Alleles and Genetic Distance																	
Generations	Years	12/0	12/1	25/0	25/1	25/2	37/0	37/1	37/2	37/3	37/4	67/0	67/1	67/2	67/3	67/4	67/5	67/6	67/7
4	100	34	7	61	27	8	83	59-83	31-46	11-31	4-12	90	71	46	12-24	10	4	1-4	0-1
8	200	56	20	85	58	29	97	90-97	71-86	45-71	27-44	99	95	86	48-69	47-49	29-33	17-30	6-15
12	300	71	33	94	78	52	100	98	90-97	74-91	58-73	100	99	97	77-91	79-81	64-68	50-65	27-46
16	400	81	46	98	89	70	100	100	97-99	90-97	80-89	100	100	100	92-98	94	86-89	77-87	55-74
20	500	87	57	99	95	83	100	100	99	96-99	92-96	100	100	100	97-100	99	96-97	92-96	78-90
24	600	91	68	100	98	90	100	100	100	99-100	98-99	100	100	100	99-100	100	99	97-99	91-97

Excellent probability of a common ancestor in genealogy terms.
Fair probability of a common ancestor in genealogy terms.
Weak probability of a common ancestor in genealogy terms.
Statistically poor probability of a common ancestor in genealogy terms.

The highlighted example above shows that for 37 markers and GD=0, there's a 97% chance of MRCA in 8 generations. For 67/0 it jumps to 99%. Do you see that the more markers tested increases your probability of finding a MRCA?

Pop Quiz #2

From the table above, in the case between kit B111 and B222 (previous page) where the genetic distance was 1, what is the probability that their MRCA is within the last 4 generations?

Kinds of Y-DNA Tests

There are several types of Y-DNA testing. Which one you choose will depend on your test goals.

Y-STR

Y-DNA **S**hort **T**andem **R**epeat (STR) testing tells about the most recent generations (1 to 45) of a male's paternal heritage. This is the most common test taken and comes in various number of markers tested (e.g., 37, 67, 111). The higher the number, the more locations are analyzed (and the higher the cost!). With STR tests, you will be given a "predicted" Haplogroup, which is a conservative estimate and should be the starting point for further testing.

Y-SNP

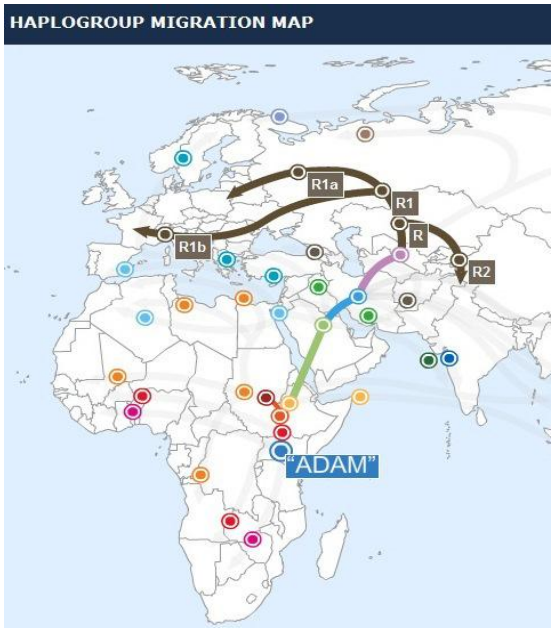
Single **N**ucleotide **P**olymorphisms (SNPs) tell of the line's deeper history and trace back to ancient times. If you test on a well researched branch of the Haplotree, you can find the origins of your patrilineal family line. SNP testing only looks at known SNPs and probably won't take you down the tree to your **Terminal SNP**.

Big-Y

The BIG Y test is a direct paternal lineage STR and SNP test. It's designed to explore deep ancestral links on our common paternal tree. It tests both thousands of known branch markers and millions of places where there may be new branch markers. It is intended for expert users with an interest in advancing science.

Single **N**ucleotide **P**olymorphisms (SNP - pronounced "snip") is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (e.g. > 1%)[3]. It tells of the heritage's deeper history and trace back to ancient times (thousands of years - millennia) and regions.

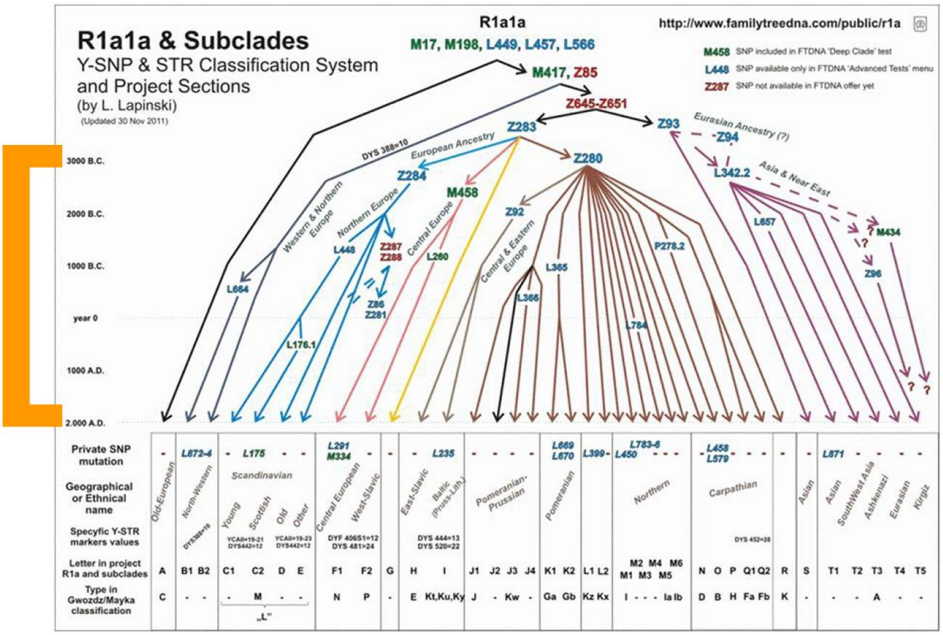
Remember that a Haplogroup is a major branch on the paternal tree who shares a common ancestor (e.g., R1b). Haplogroups are associated with early human migrations, as indicated by the map below.



R Haplotype Migration from "Y-Adam"

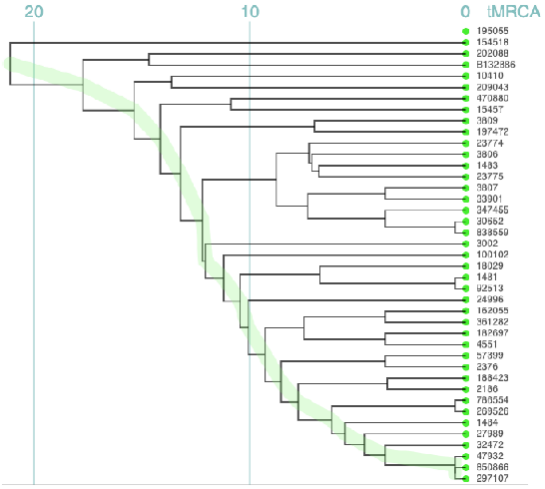
[3] FT-DNA must have at least TWO people with the same SNP to create a new terminal SNP

Today these can be associated with a geographic region or regions. The Y Chromosome Consortium (YCC) developed a system of naming major Y-DNA Haplogroups with the capital letters **A** through **T**, with further **Subclades** (subgroup of a Haplogroup) named using numbers and lower case letters (YCC longhand nomenclature). In the example, **R-M269**; R is the Haplogroup and M269 is the subclade. YCC shorthand[4] would call this R-M269. Notice the timeline along the left shows the evolution of these subclades over five thousand years. It shows the migration and isolation of the subclade to specific geographic area (e.g., European, Scottish, etc.).

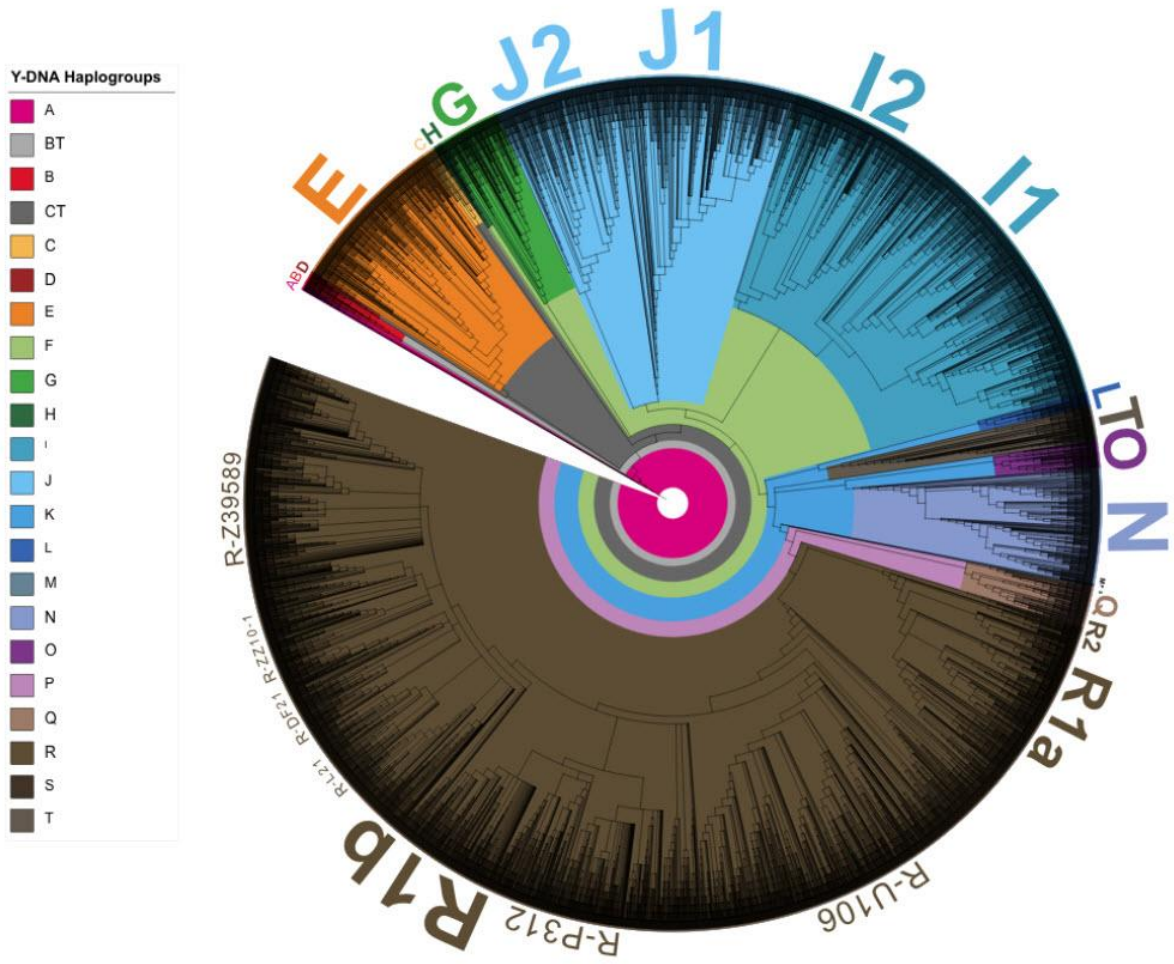


Haplogroup

The Y-DNA STR test will also give you your "predicted" Y-Haplogroup. A Haplogroup is a set of similar **haplotypes** that share a common ancestor having the same single nucleotide polymorphism (SNP) mutation in all haplotypes. The Haplogroup is shown as an alphanumeric code: e.g., R1a1a (old terminology) or R-M512. In the chart below, each "branch" is a mutation for that family that starts a unique Haplogroup for those men and the "branches" under are new subclades to that Haplogroup. So the 1st (top) kit in the group is "*Related*" to the last (bottom) kit, but it is 21 generations back!



[4] https://en.wikipedia.org/wiki/Conversion_table_for_Y_chromosome_haplogroups



Y-Haplogroups at FTDNA

Matching

When you get your results, you will be in one (or more) of the following categories:

1. Exact match with one or more persons (Genetic distance = 0)
2. Close match with one or more persons (Genetic distance = 1-3)
3. Distant match with one or more persons (Genetic distance = 4-?)
4. No matches at 37 (or more) markers

We will cover what each category means and what to do in each one.

Exact Match

This is good! An exact match **verifies** your paternal relationship with that person. The only thing to do here is be happy! You hopefully have met your objective.

Close Match

A close match means that you are most likely cousins or that you may have had mutations[5] in your family on the fast changing STR markers. So if you have a close match, contact that person to see who the common ancestor may be (if you don't already know). Ask them for surnames (or a **GEDCOM**) and add them to your tree and find the relationship yourself!

Distant Match

A distance match may mean that you are related farther up the tree than you have researched or discovered. You may have luck at mitoYDNA.org (covered in Lesson#3 - Tools) extending your tree with a match there. If not, then you will have to work with your distant match contact to find the relationship together (triangulate).

No Match

If you have absolutely no matches for your surname, it could be the result of one of the following:

1. No one from your surname has tested at FTDNA. Or they only tested at 25-markers (or less). Wait for (or FIND!) someone else with your surname to also test. Try looking at mitoYDNA.org for matches.
2. An unusual last name that few/none have tested. Use mitoYDNA.org to see if someone tested at another company.
3. There was a surname change in your family in the past. Look for variations in spelling. Again, use mitoYDNA.org to see if someone tested at another company.
4. **N.P.E./MPE** - Non-Paternal Event / Not Parent Expected -or- Misattributed Parentage Event
5. You didn't meet the threshold to have the match listed. FT-DNA has established match thresholds. For someone to match, they have to meet the criteria[6] for Genetic Distance in the following table:

[5] Bennett Greenspan, CEO of FT-DNA has a GD=1 with his father!!
[6] FamilyTree DNA

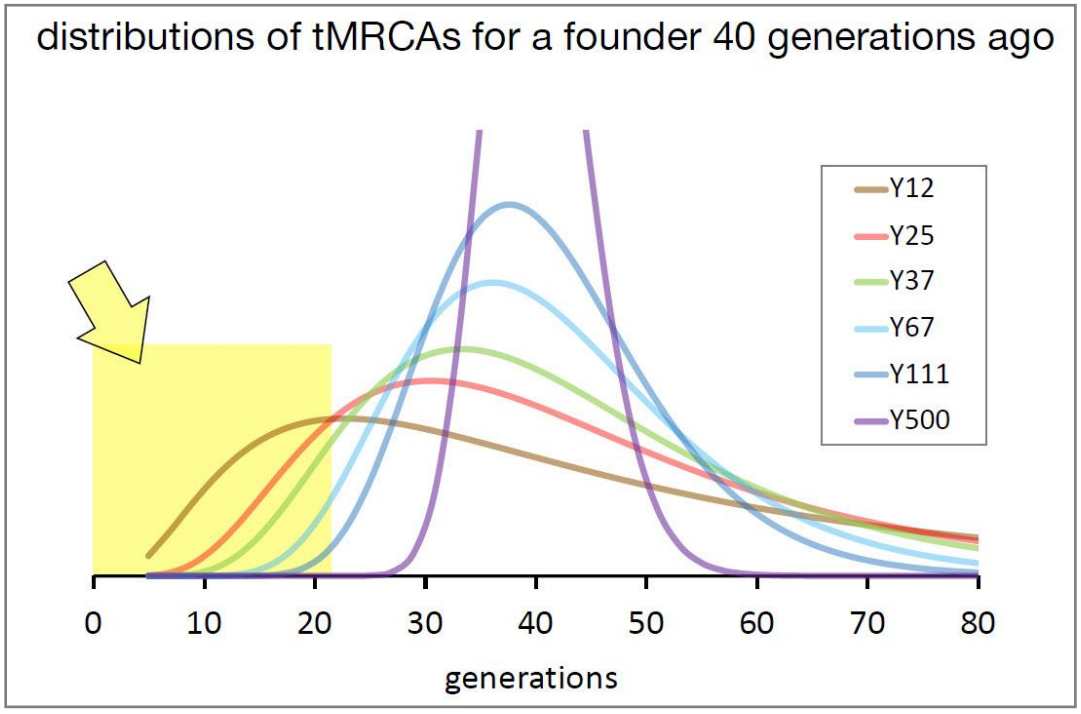
# Markers Tested	Maximum # Mutations Allowed
12	0[7]
25	2
37	4
67	7
111	10

One of the things you might try if you tested at 37-markers (or lower) is to upgrade your test to 67-markers. That might show some matches at a greater distance (5-6) which didn't show up at 37-markers.

In **all** of the 'No Match' cases, you may also want to do an autosomal test and then concentrate on the matches that you know are on your paternal side (by genealogy or Not In Common With (ICW) maternal matches). You may find "cousin" matches that help you go 'up' the tree and back down to your paternal match. It may take a combination of bouncing back and forth between Y-DNA searching and Autosomal searching to find your paternal ancestor(s).

One interesting fact that I got from the FTDNA Project Manger's Conference in 2019 came from Rob Spencer's presentation on the "Big picture" analysis of Y-DNA. There is a Rule of Thumb: for any set of ***N*** men, you will find **0.6***_N_** Surname lineages; that means:

- 1/3 of men will have **NO** matches

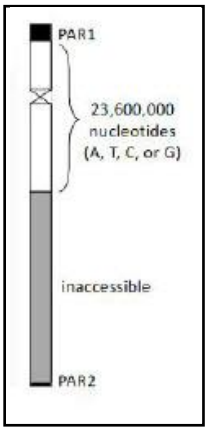


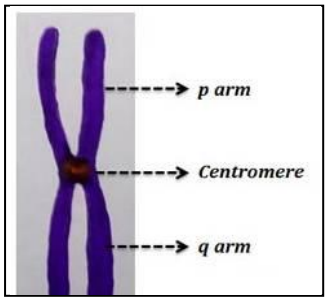
[7] If while in the FT-DNA match list, you go from higher markers (37 or 67) down to 12 markers, you may see it show a GD=1. This is a remnant of showing the higher markers. It won't initially show a 12 marker match >0. Or you may see GD=1 if your match is in the same Project.

Section 2. Y-DNA Details

Y-DNA Composition

In the nucleus of each cell, the DNA molecule is packaged into thread-like structures called chromosomes. Each chromosome is made up of DNA tightly coiled many times around special proteins. The entire Y chromosome (right) has roughly 57M nucleotides, or base pairs, of which about 23M are accessible. The genealogically relevant regions (**white**) are those that are passed intact from father to son with high fidelity. The two tips of the Y chromosome are called the **telomeres**; these regions are subject to recombination with the X chromosome (**black**; **P**seudo **A**utosomal **R**egions PAR1 and PAR2) and don't have a stable ancestral reference genome value so therefore they are of limited use for genealogy applications. The other region (grey) is a highly repetitive sequence, and therefore inaccessible to NGS sequencing (inexpensively).



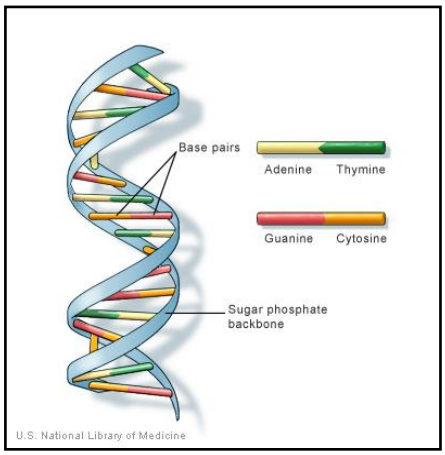


Each chromosome has a constriction point called the centromere, which divides the chromosome into two sections, or "arms". The Y chromosome is made up of two segments which are named the "**p**" arm ('p' stands for "petit") and the "**q**" arm (merely because '**q**' is the next letter). The arms are joined at the **centromere** which is usually shown as a narrow "neck".

Base Pairs

The Y chromosome is made up of strands of DNA and the "units" of DNA are nucleotides that contain four different bases:

1. Adenine (A)
2. Thymine (T)
3. Guanine (G)
4. Cytosine (C)



The Y chromosome can also be considered as one *very* long string of these values (e.g., AGTCGATA…). Because DNA is a double helix, each base is bonded to another on the other strand and forms what's called a base pair.

The **human genome** has been mapped and the reference in use across *most* of genetic genealogy today is Genome Reference Consortium Human Build 38, usually abbreviated as GRCh38 or **HG38**. As of Dec. 2013 it is the standard reference[8] assembly sequence used by **N**ational **C**enter for **B**iotechnology **I**nformation (NCBI).

[8] Wikipedia - https://en.m.wikipedia.org/wiki/Reference_genome

In HG38 the Y chromosome is 57,227,415 base pairs in length, and these are numbered starting with 0 at one end and 57,227,415 at the other.

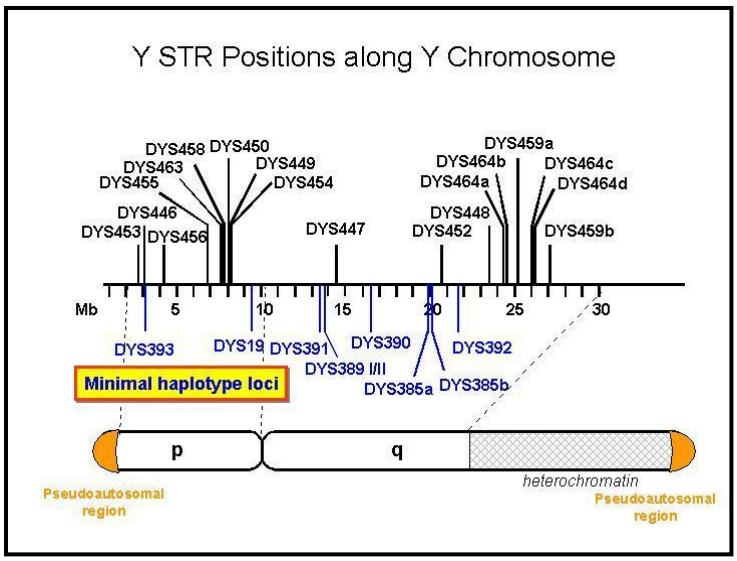
If your research shows that a SNP has a position of 12345678, that means it's at a base pair location on the Y chromosome found by counting 12,345,678 base pairs over from the origin. SNP testing and BigY results will refer to these locations.

DYS Prefixes and Numbers

The prefixes used for STR markers found by Y-DNA testing have special meaning. STR markers are named according to guidelines published by HUGO[9] Gene nomenclature committee. They are:

- **D** - DNA
- **Y** - Y chromosome
- *n* - for the complexity of the repeat segments:
  - **S** - a unique segment
  - **Z** - a number of repetitive segments at one site
  - F - a segment that has multiple copies[10] on the Y
- FTY - stands for Family Tree Y (acts as a place-holder until HUGO assigns an official prefix

All STRs are given a unique identification number (after its prefix). For example DYS393 indicates that it is a **D**NA **Y** chromosome [**S**] unique segment with 393 as its identifier.



c/o - FTDNA

[9] Human Genome Organisation (HUGO)
[10] See "*Palindromes*" on Page 17

Mutations

The problem with STRs is that they can go up or down at random and it is often not possible to tell the order in which changes occurred. Co-incidental changes and "back-mutations" (where a number changes back to an earlier state) can confound the interpretation of results.

SNPs, by contrast, tend to be "one-off" events. They still occur at random but we can generally tell the order in which they occurred. They can be very helpful for identifying individual family lines.

When mutations take place, they come in three major types:

- a **Deletion** - when nucleotide(s) fail to copy during reproduction. DNA data at that location is permanently missing
- a **Replacement** - when a nucleotide is replaced; it's where a **G** mutated to a **T** or a **C** to an **A**.
- an **Insertion** - when nucleotide(s) duplicate and are inserted between existing nucleotides.

Mutation Rates

Knowledge of Y-chromosomal short tandem repeat (Y-STR) mutation rates[11] is essential to determine the most recent common ancestor (MRCA) in familial searching or genealogy research. The following table shows estimates or rates (e.g., GD=1 at 12 markers means MRCA is 40-120 generations -or- that Y-DNA mutates much faster that mtDNA). FTDNA has 13 "*fast moving*" markers based mutation rate deductions and the presumption that "*fast moving*" means faster than .0028.

Basic DNA Tests			
Rates & MRCA		**GD=0**	
Y-DNA 12 marker	Mutation Rate - .0019	Y-DNA 12 marker	16 to 80 generations
mtDNA HVR1	Mutation Rate - .00003	Y-DNA 25 marker	6 to 32 generations
Y-DNA 12 marker	16 to 80 generations	Y-DNA 37 marker	3 to 15 generations
Y-DNA 12 with GD=1	40 to 120 generations	Y-DNA 67 marker	2 to 10 generations
Y-DNA 12 with DYS426 mismatch	240 to 800 generations	Y-DNA 111 marker	1 to 6 generations

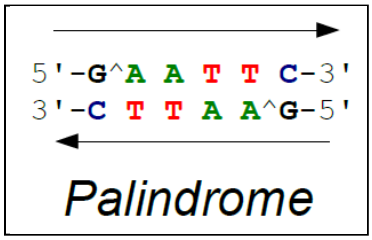
Below is a mutation rate table developed by J. D. McDonald and used in his mutation calculator tool. Tools will be covered in Lesson #3.

Y-DNA STRs		**Mitochondrial, Autosomal & SNPs**	
FTDNA Y-12 STR	0.0020	mtDNA complete genome	3·10-6 (type 3.e-6)
FTDNA Y-25 STR	0.0026	mtDNA coding	2·10-6 (type 2.e-6)
FTDNA Y-37 STR	0.0042	mtDNA HVR	2·10-5 (type 2.e-5)
FTDNA Y-67 STR	0.0029	autosomal and X SNPs	1·10-8 (type 1.e-8)
FTDNA Y-111 STR	0.0026	Y SNPs (BigY etc.)	3.0·10-8 (type 3.e-8)

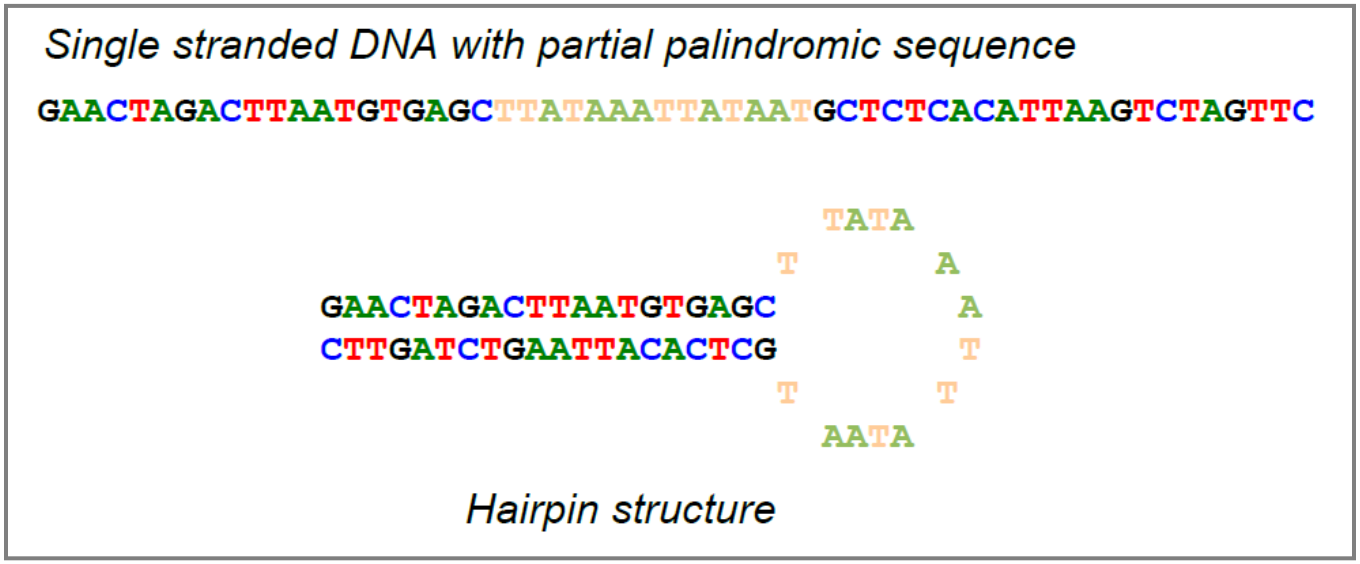
[11] https://en.wikipedia.org/wiki/List_of_Y-STR_markers

Combined with Genetic Distance, formulas can predict the **TMRCA** (**T**ime to **M**ost **R**ecent **C**ommon **A**ncestor). For the Y chromosome these rates assume a **31** year generation.

Palindromes



From Wikipedia: "A palindromic[12] sequence is a nucleic acid sequence in a double-stranded DNA molecule wherein reading in a certain direction on one strand matches the sequence reading in the same direction on the complementary strand. A palindromic nucleotide sequence is capable of forming a hairpin" (see below). Multi-copy STRs are located on the symmetrical arms of palindromes. The sequences on the palindromic arms are nearly identical and are highly prone to recombination mutations. These mutations cause Genetic Distance mismatches in closely related individuals. Differences in allele values for markers like DYS464 are affected.



c/o - FTDNA

Pop Quiz #3

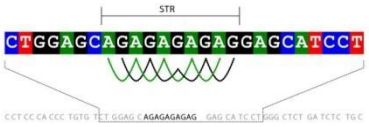
How many base pairs are there in the Y Chromosome (HG38)?

[12] Palindrome - a word or phrase that reads the same forwards or backwards: 'radar', 'madam I'm Adam'

Section 3. STRs vs SNPs

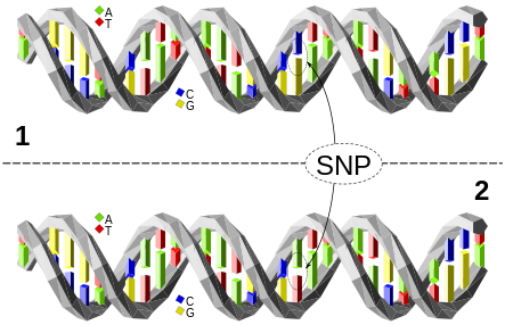
STRs

Short **T**andem **R**epeat (STR) testing is based on repetitions of nucleotides. It reveals information on the most recent generations (1 to 21) of a male's paternal line. It gives you a match to a surname.



SNPs

Single **N**ucleotide **P**olymorphisms (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (e.g. > 1%)[13]. In the example chart, it's where a **G** mutated to a **T** and a **C** to an **A**. It tells of the heritage's deeper history and trace back to ancient times (thousands of years - millennia) and regions. Just like STRs, SNPs are passed intact from father to son with high fidelity.



When speaking of SNP ages, the terms 'upstream' and 'downstream' are used to show age relationship. Upstream means older (SNPs, haplogroups, branches, etc) within the same branch; for example R1a is upstream of CTS3402. Downstream means younger within the same branch; for example CTS3402 is downstream of R1a.

SNP Age

Normally people try to get the closest applicable years-per-SNP average (often "144" or "83" are quoted) and multiply the number of SNPs under a particular branch to arrive at an age estimation for the SNP at the top of that branch. Tools for SNP Age calculation will be covered in Lesson #3.

Phyloequivalents

Many SNPs were independently discovered more than once and are listed with multiple names. This is called **phyloequivalents**. Most haplogroups have multiple phyloequivalent SNPs. The SNP name starting letter indicates the person/organization that found this SNP (e.g., CTS = Chris Tyler Smith, BY = BigY, FT=FTDNA). **R1a** has over 100 phyloequivalent SNPs!

R1a	**L146/M420/PF6229**, L62/M513/PF6200, L63/M511/PF6203, L145/M449/PF6175
R1a~	CTS903/M610/PF6154, CTS2443/PF6178, CTS2907/M665, CTS3877/PF6184, CTS4509/M687, CTS5273/PF6190, CTS5936/M698/PF6192, CTS6918/PF6196, CTS7559, CTS8008/M726, CTS8851/M740, CTS9515/M744, CTS9596/M745/PF6205/V3655, CTS10627/M786, CTS11148/M796, CTS11530, CTS11734/M800/PF6226, CTS12276, CTS12639/M814/PF7535, F886/PF6153, F928/M616, F1088/M629/PF6160, F1769/M662/PF6179, F2948/M752/PF7527/V3820, F3364/M794/PF6222, F3466/M803/PF7534, F3570/PF6233, F3650, FGC32014/Y215, FGC32015/Y209, FGC32438/Y217, L457/PF6191, M641/PF7516, M767/PF6212, M768/PF6213, M775/PF6215, M784/Y216, PF6189

Pop Quiz #4

In addition to mutation rate, what is the other factor used to calculate the **T**ime to **M**ost **R**ecent **C**ommon **A**ncestor (TMRCA)?

[13] FT-DNA must have at least TWO people with the same SNP to create a new terminal SNP

Section 5. Glossary

- **Allele** - One of the possible values for an STR marker.
- **Autosomal** - DNA test that is designed to find relatives on any of your ancestral lines within the last 5-6 generations
- **Base Pair** - A base pair is a unit or building block of DNA. Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) are the four primary bases in DNA. The order of bases is the sequence of DNA.
- **Centromere** - the specialized DNA sequence of a chromosome that links a pair of sister chromatids (exact copy of ½ chromosome)
- **Genetic Distance** - Genetic Distance is the number of differences, or mutations, between two sets of results. A genetic distance of zero means there are no differences in the results being compared against one another
- **Haplogroup** - A Haplogroup is a major branch on either the maternal or paternal tree of humankind. Haplogroups are associated with early human migrations. Today these can associated with a geographic region or regions.
- **Haplotype** - A haplotype is the set of allele values for the results of DNA test. For example, the results of the Y-DNA12 test for one person is their haplotype.
- **HG38** - the assembly of the human genome that are the result of genome sequencing and analysis projects
- **Locus** - A locus is a specific location in your genetic code. In a genetic map of our DNA, the locus tells us where to find any base. The plural of locus is loci.
- **Marker** - A marker is a physical location (locus) on the chromosome. The term is often used colloquially in genetic genealogy to refer to a short tandem repeat (STR).
- **Mitochondrial** - The DNA line that follows your mother's maternal ancestry
- **Mutation** - A random naturally occurring change in the chromosome
- **Phyloequivalents** - similar SNP(s) discovered independently, but it is unclear of the order in which they occurred. More testing may show the true order of mutations.
- **Short Tandem Repeat (STR)** - A short DNA motif (pattern) repeated in tandem. ATGC repeated eleven times would give the marker a value or allele of 11
- **SNP** - **S**ingle **N**ucleotide **P**olymorphisms
- **Subclades** - The a subgroup of a Haplogroup
- **Terminal SNP** - The SNP furthest down the Haplotree for which you have tested.
- **Telomere** - a cap of DNA that protects the rest of the chromosome from damage
- **TMRCA** - **T**ime to **M**ost **R**ecent **C**ommon **A**ncestor, the probability that two people have a certain number of generations between them and a common ancestor.

- **X-Chromosome** - Is reported as part (Chromosome 23) of the Autosomal test. A man passes down his X chromosome only to his daughter. A woman passes down her X chromosome to both her sons and daughters.
- **Y-Chromosome** - One of the two sex chromosomes, X and Y. The Y-Chromosome passes down from father to son. Females do not receive it. As the Y-Chromosome is passed on through the paternal line, it is valuable for surname based genealogy studies.
- **Y-DNA Backbone test** - If a person's Y-DNA haplogroup cannot be predicted with 100% confidence, the SNP Assurance Program at FTDNA will test your sample with our Backbone SNP test for FREE. This test is a "deep" multiple SNP test. Specifically, if we cannot predict a person's Y-DNA haplogroup with sufficient confidence that they can join the National Geographic's Genographic Project, we will automatically perform a Backbone SNP test in order to identify the haplogroup assignment. Backbone tests take about 6-8 weeks from the time they are ordered.

http://www.dnaadoption.com/index.php?page=acronyms
http://www.dnaadoption.com/index.php?page=glossary-of-dna-terms

Section 6. Resources

Once you have clues to people who might match your DNA, there are gold mines of resources just waiting to help you move farther along in your search.

Internet

- The resources we've used in class, of course: FamilyTreeDNA.com
  https://www.familytreedna.com/learn/ftdna/webinars
- DNAAdoption - www.dnaadoption.com
- ISOGG: http://www.isogg.org
- BYU: https://www.youtube.com/watch?v=MP3z3hLLDog **Basic Y-DNA Webinar [50:57]**
- NCBI - https://www.ncbi.nlm.nih.gov/
- GRC - https://www.ncbi.nlm.nih.gov/grc
  - Home - genomereference.org
  - Presentations - https://www.slideshare.net/GenomeRef/presentations
- Mutation Rate
  - Tim Jantzen's mutation spreadsheet **Spreadsheet**
  - https://en.wikipedia.org/wiki/List_of_Y-STR_markers
  - http://www.kerchner.com/dnamutationrates.htm
- DNAeXplained - https://dna-explained.com/2016/06/29/concepts-genetic-distance/
  https://dna-explained.com/2016/07/27/y-dna-match-changes-at-family-tree-dna-affect-genetic-distance/
  https://dna-explained.com/?s=Y+DNA+Resources+and+Repository
- There are many surname project websites, often associated with Y-DNA tests. https://learn.familytreedna.com/group-projects/family-tree-dna-group-projects-charge-money-run-belong-one/

Books

- Aulicino, Emily - "***Genetic Genealogy: The Basics and Beyond***", ©2014, AuthorHouse LLC, Bloomington, IN [ISBN-13: 9781491840900]
- Bettinger, Blaine T. - "***Guide to DNA Testing and Genetic Genealogy***", ©2017, Cincinnati, Ohio : Family Tree Books
- Smolenyak, Megan & Turner, Ann - "***Trace Your Roots With DNA: Use Your DNA to Complete Your Family Tree***", ©2004, Rodel Books, New York [ISBN 1594860068 (ISBN13: 9781594860065)]
- Pomery, Chris - "***DNA and family history : how genetic testing can advance your genealogical research***", ©2004, Dundurn Group, Toronto, Ontario
- Dowell, David R. (Ph.D.) - "***NextGen Genealogy: The DNA Connection***", ©2015, Santa Barbara, California : LIBRARIES UNLIMITED, an imprint of ABC-CLIO, LLC
- Hill, Richard - "***Finding family : my search for roots and the secrets in my DNA***", ©2012, Richard Hill (self published), Grand Rapids, Michigan. *<< available as book or e-Book>>*
- Griffeth, Bill - "***The Stranger in My Genes***", ©2016, UPNE Book Partners, Lebanon, NH [ISBN-13: 978-0-88082-344-9]
- Vance, J. David - "***The Genealogist's Guide to Y-DNA Testing for Genetic Genealogy***". ©2020, Amazon Publishing, [ISBN: 9798621504779] *<< available as book or e-Book>>*

You

- Use your personal network. If you have friends or family who are interested in family history or genealogy, brainstorm with them about where to look for information. They may have ideas or resources that can help in your search.
- Give back. Share what you know and help others when you can. The old saying, "What goes around, comes around" has been proven over and over and over.

Section 7. Pop Quiz Answers

1. Genetic Distance = **4**! Though the difference is on a single DYS (385), since both values are different that counts as 2. So with DYS 390 &389I2 also being different, the total is four.

Kit	Surname	DYS393	DYS390	DYS19	DYS391	DYS385	DYS426	DYS388	DYS439	DYS389I	DYS392	DYS389II
B111	Smith	12	23	14	10	17-18	11	12	11	12	14	**29**
B222	Smith	12	23	14	10	17-18	11	12	11	12	14	**28**
B333	Smythe	12	22	14	10	**16**-17	11	12	11	12	14	29

2. 7% - It's a 12 marker test with GD=1; you'd have a 68% chance that it's 24 generations (impossible)!

Generations	Years	12/0	12/1
4	100	34	7
8	200	56	20
12	300	71	33
16	400	81	46
20	500	87	57
24	600	91	68

3. 57,227,415 ("*Y-DNA Composition*", Page 12)

4. Genetic Distance. ("*Mutation Rates*", Page 14)

Section 8. Epilogue

Class Notes

Summary

Y-DNA can trace the origins of a surname and can be used to find other people who shared an ancestor on this paternal line.

Access to Materials

Your Moodle login will continue to be available to you. If you'd like to retake the class, you are welcome to sign up as space is available. No additional donation will be requested (although they are certainly welcome!). Email moodle@dnaadoption.com if you wish to re-take this class.

Class Forum

The class forum will be available for follow-up questions for an additional 4 weeks subsequent to the posting of Lesson 3. Participants will also have the option of re-taking this class for free within the next year. Contact moodle@dnaadoption.com if you wish to re-take this class.

Stay in Touch! Let us know how your journey is going.

moodle@dnaadoption.com