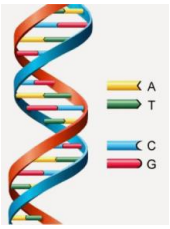
STR vs. SNP

Gale French - DNA Adoption

There is often confusion in Y-DNA testing on the differences between **S**hort **T**andem **R**epeat (**STR**) tests and **S**ingle **N**ucleotide **P**olymorphism (**SNP**) tests and their meanings and uses. They are quite different, but both have their purpose in genealogy and parental searches.

The DNA helix consists of base pairs that attach to each other in a certain sequence to reproduce the DNA of a parent(s). These pairs are made up of:

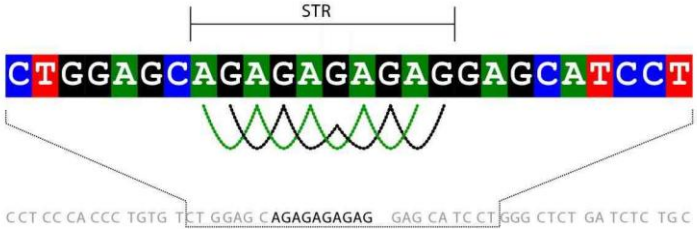


1. Adenine (A)
2. Thymine (T)
3. Guanine (G)
4. Cytosine (C)

T usually combines with **A** and **C** typically combines with **G**.

STR

When tested, the sequence of these pairs is determined and certain series of repeated pairs appear and are counted at that location (a Marker). Each marker has a DYS name (e.g., DYS-385a). In the example below, the pair of AG is repeated 4 times after the initial pair so the repeat count (the allele value) is 5.



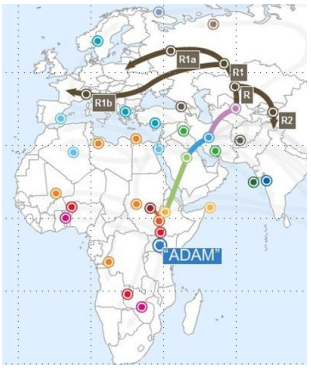
This is a **S**hort **T**andem **R**epeat**.** STR Tests today come in 12, 25, 37, 67 and 111 markers tested. The test results show who you have matched and if there were any difference in the marker values (the **G**enetic **D**istance or GD). The GD tells you how "much" you are related and charts and tools can give you an estimate of how many generations back you might find the Most Recent Common Ancestor (MRCA). This test will also give you your Haplogroup; the major branch of the tree who share a common ancestor (e.g., R1b). A subclade is a subgroup of a Haplogroup. It will also show you the Surnames of your matches and (if available) their 'paper' research of their most distant ancestor. So this is "recent" (or genealogical) history.

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

SNP

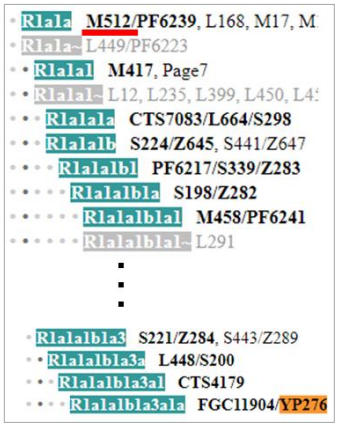
A SNP is a DNA variation in a single nucleotide that occurs at a specific position in the genome (locus), where each variation is present to some appreciable degree within a population group (e.g. > 1%). This is where a G might have mutated to become a T and/or a C may have become an A.

For example, if you tested 8 individuals and had the following results:




```
Individual 1: AAGGTG C AGCAGTC
Individual 2: AAGGTG T AGCAGTC
Individual 3: AAGGTG T AGCAGTC
Individual 4: AAGGTG T AGCAGTC
Individual 5: AAGGTG C AGCAGTC
Individual 6: AAGGTG C AGCAGTC
Individual 7: AAGGTG T AGCAGTC
Individual 8: AAGGTG T AGCAGTC
```

The position highlighted in bold is a SNP. These mutations (polymorphisms) can be traced back thousands of years. So SNP testing represents historical (or millennial) history. Subclades are defined by a "terminal" SNP; the SNP furthest down in the Y chromosome tree. In the chart to the right, a simple R1a Haplogroup can be "refined" to a Haplogroup subclade of YP276; which can sometimes help to determine if a match is closely related by both testers having the same terminal SNP. This "refinement" can come with a SNP Pack test that tests about 100 SNPs (for around $100) or through the Big-Y test which looks at about 11 million SNPs (for about $400 to $600). A Big-Y test will also give you a list of matches who have your SNPs as well (a SNP Pack only gives you a terminal SNP). If you are trying to sort out a match and only want a terminal SNP, then a SNP pack is a cost effective method of obtaining that.



- **Short Tandem Repeats (STRs)**
  - Recent history (surname)
- **SNPs (Single Nucleotide Polymorphism)**
  - 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%)
  - where G➜T & C➜A
  - Ancient History (millennia)