This page has been reformatted to facilitate upload to our website www.S16264.info
Note that the embedded links may not work in this .pdf version

R1b-S16264: Information for Newcomers

This information is divided into two parts:

STRs, SNPs and Phylogenetic Trees (the previous page)
and
Y-DNA STR Tests (this page)

Y-DNA STR Tests

Prepared by Nigel Bond

Version of 10 March 2017 – Check here for updates

Many men start their genetic genealogy research with a Y-DNA STR test such as Family Tree DNA's Y37 test. This test determines the tested sample's **allele counts** of 37 STR markers with **mutation rates** ranging from very slow to very fast. STRs with very slow mutation rates have a very high probability of remaining unchanged over many generations, equivalent to thousands of years. Those with very fast mutation rates have a significant probability of mutating once every few generations. Others in the test have mutation rates between these extremes. The 37 STRs are selected to provide evidence of **ancient ancestry** via the largely unchanging very slow mutation rate STRs and of more **recent paternal relationships** via the more frequently changing fast mutation rate STRs. With 37 STRs there is a good probability that one or more of these STRs will have mutated say every 3 to 5 generations. However this is not inevitable and STR haplotypes can sometimes remain unchanged over many generations. On the other hand each STR mutation is a **random event** which can occur at any time: a son's haplotype may differ from his father's by 1 or even 2 mutations. This randomness introduces some uncertainty in the interpretation of STR test results. Their interpretation is further complicated by the fact that mutations can occur in either direction, increasing or decreasing the allele count at a marker location. Two apparently identical haplotypes may conceal a forward mutation followed by a **back mutation** to the original allele count. Two STR haplotypes may also be near identical due to **convergence** in which, by chance, similar mutations have occurred in different paternal family lines (probably at different times) resulting in a false apparent close relationship. The probability of false matches and hidden relationships is reduced by testing more markers: 67 markers is the minimum recommended by many experienced genetic genealogists and some advocate 111 markers for all.

How then do we interpret the results of STR tests?

The first important consideration is that the majority of the STRs will most likely have remained unchanged for thousands of years, particularly the slow and very slow mutation rate STRs, although they will probably have mutated within the last 10,000 to 30,000 years. This distant timeframe is the time by which the major haplogroups had formed and so each of the ancient haplogroups, including S16264's ancestral R1b, has a clear **STR haplotype signature** carried by the overwhelming majority of men within that haplogroup. **R1b-M269+** men will generally have all or most of the following allele counts (FTDNA marker numbers are listed in parentheses, the allele counts in square brackets) on these very slow mutation rate STRs:

(7) DYS426 [12]; (8) DYS388 [12]; (11) DYS392 [13 or more]; (17) DYS454 [11]; (37) DYS438 [12]; (42) DYS590 [8]; (45) DYS472 [8].

In addition most men in the **P312+** clade of M269+ (which includes S16264+) will have **(66) DYS492 [12]**. The ancestor of all these P312+ men lived approximately 5000-4700 years ago. The ancestor of all L21+ men lived approximately 4800-4500 years ago and the ancestor of all S16264+ men lived approximately 4200-3800 years ago. *(All ages to be confirmed by on-going research)*

The next important consideration is that, by statistical analysis of known P312+ men's haplotypes, we can define the most frequently occurring allele counts at each of the DYS marker locations for the P312 population. These are the **P312 modal** allele counts. Listed in the standard FTDNA sequence they form the **P312 modal haplotype** (sometimes called the **Western Atlantic Modal Haplotype** or **WAMH)**:

Y37 P312 Modal -
13 24 14 11 11-14 12 12 12 13 13 29 17 9-10 11 11 25 15 19 29 15-15-17-17 11 11 19-23 16 15 18 17 36-38 12 12

Marker 38-67 P312 Modals -
11 9 15-16 8 10 10 8 10 10 12 23-23 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12

and Marker 68-111 P312 Modals are -
33 15 9 16 12 26 26 19 12 11 13 12 10 9 12 12 10 11 11 30 12 13 24 13 10 10 20 15 19 13 24 17 12 15 24 12 23 18 10 14 17 9 12 11

Having confirmed that a haplotype is almost certainly P312+ the next step is to identify what differentiates it from most other P312s. That is achieved by identifying the markers which are different from those in the P312 modal haplotype: these are its **off-modal** markers. If more than 10% of the markers are off-modal then these may be sufficient to identify related haplotypes within the tested population. The most useful sources of haplotypes for comparison are Mike Walsh's Y67 and Y111 Haplotype worksheets which are available to members of the FTDNA R1b-M343 and R1b-L21/S145 Projects. The most significant off-modals for matching are those which occur at relatively low frequencies in the tested population: in say 10% or less or, better, in say 3% or less *(see the Allele Frequency file in the S16264 Project files folder for frequency data for our R1b-L21 clade)*. Sharing just one of these is not sufficient to propose a shared paternal ancestor within the last 3000 years. Instead several are needed.

Using my Y67 haplotype (FTDNA 170191) as an example, highlighting my off-modals in **red**

12 24 15 10 11-14 12 12 13 13 13 29 16 9-10 10 11 25 15 19 29 14-15-15-15 11 11 19-23 16 15 18 16 36-36 12 12 11 9 15-16 8 10 10 8 10 10 12 23-23 16 10 12 12 15 8 14 22 21 14 12 11 13 11 11 12 12

There are 14 off-modals at DYS 393 [12]; 19/394 [15]; 391 [10]; 439 [13]; 458 [16]; 455 [10]; 464a [14]; 464c [15]; 464d [15]; 570 [16]; CDYb [36]; 444 [14]; 520 [21]; 446 [14]. This is a relatively large number of off-modals and is therefore likely to provide a good basis for finding meaningful matches with other men's haplotypes. The frequency at which these allele counts occur within L21 are 393 [12] 2%; 19/394 [15] 7%; 391 [10] 32%; 439 [13] 13%; 458 [16] 16%; 455 [10] 1%; 464a [14] 11%; 464c [15] 8%; 464d [15] 1%; 570 [16] 15%; CDYb [36] 6%; 444 [14] 1%; 520 [21] 9%; 446 [14] 16%. Several of these occur at frequencies of less than 10%: a match sharing several of

these is likely to have a shared paternal ancestor living within 3000 years before the present day. The greater the number of matches the more recent will be the shared paternal ancestor.

As has already been mentioned, the marker mutation rate is an important consideration: an off-modal of a slow mutation rate marker is more likely to remain unchanged over many generations than an off-modal of a faster mutation rate marker (although any off-modal can occur at any time – the slow mutation rate off-modal may be more recent than the faster mutation rate off-modal). Broadly, the markers can be categorised as follows (from Mike Walsh's Y67 database):

Very Slow – DYS 426, 388, 455, 454, 578, 641, 472,425, 436, 490, 450

Slow – DYS 393, 392, 437, 438, 531, 395S1a&b, 590, 537, 594, 617, 568, 487, 640, 492, 565

Medium – DYS 390, 19/394, 391, 385a&b, 389i&ii, 459a&b, 447, 448, GATA-H4, YCAIIa&b, 442, 406S1, 511, 413a&b, 557, 444, 520, 572

Fast – DYS 439, 458, 449, 464a-d, 460, 456, 607, 576, 570, 534, 481, 446

Very Fast – DYS CDYa&b (so fast that mismatches can be considered insignificant)

In my haplotype 455 [10] can be considered particularly significant as it is both low frequency and very slow: if the mutation occurred many generations ago it is likely to be shared by a substantial number of descendant men living today. This has proven to be the case as we have evidence that my 455 [10] off-modal originated over 1000 years ago as shown by my distant cousin Sylvain Leprovost (FTDNA 162326). His Y67 haplotype is:

13 24 16 10 12-14 12 12 12 13 13 29 16 9-9 10 11 25 15 19 29 15-15-15-15 13 11 19-23 15 15 17 16 35-39 12 12 11 9 15-16 8 10 10 8 10 10 12 23-25 16 10 12 12 15 8 14 22 21 13 12 11 13 11 11 12 12

Sylvain and I share 8 off-modals (underlined above) at DYS 391 [10]; 458 [16]; 455 [10]; 464c [15]; 464d [15]; 570 [16]; 444 [14]; 520 [21]. In addition, we both have DYS 19/394 greater than P312 modal [14] with allele counts of [15] and [16]. These similarities are sufficient to claim a shared common ancestor who had these same off-modals. This has now been proven by SNP tests as we are both S16264+. We have also found another group of men who share this ancestry. They descend from a man who migrated from Britain to America in the seventeenth century *(see S16264 Project Files – 'Norman 45510 – An Anglo-Norman cluster within L21' for more details).*

Our shared ancestor with this distinct haplotype lived approximately 2000 years after the birth of the first S16264 man. There will many other S16264+ men who have haplotypes with very different off-modals than Sylvain and myself. Our S16264+ Project will collect the haplotype data and use them to determine the relationship between the various S16264+ families.

How do we measure the closeness of relationship between STR haplotypes?

The difference between haplotypes is quantified by **Genetic Distance** (**GD**). This is calculated in two ways:

The most commonly used is the **1-step Genetic Distance.** This is the sum of the absolute differences in allele count at each marker for the two haplotypes. **GD37** is the 1-step Genetic Distance for FTDNA's standard 37 markers, **GD67** for FTDNA's standard 67 markers, **GD111** for FTDNA's standard 111 markers, **GD27** is for the 27 slowest mutating markers within the standard

67 markers, **GD35** is GD37 excluding fast mutating CDYa&b and **GD65** is GD67 excluding fast mutating CDYa&b. For example two of these 1-step GDs between myself and Sylvain Leprovost are GD37=14 and GD67=17. Identical haplotypes have all GDs equal to zero. The larger the GD the more distant is the relationship.

Sometimes researchers use the **Infinite Allele Genetic Distance**. This simply counts the number of markers which have different allele counts regardless of the magnitude of the difference. For example the 67 marker infinite allele GD between myself and Sylvain is 13.

Genetic Distance provides only an approximate indication of the probable degree of relatedness between two haplotypes due to the random nature of STR mutations. The degree of relatedness is usually expressed in terms of 'the number of generations to the most recent common ancestor': for Y-DNA haplotypes this being the most recent shared paternal line ancestor. This **time to most recent common ancestor (TMRCA)** can be calculated using statistical methods which are too complicated to describe here (and for me to fully understand!). Two widely used methods are:

Tim Janzen's method at http://www.timjanzen.com/dna.html

Dean McGee's method at http://www.mymcgee.com/tools/yutility.html?mode=ftdna_mode

Y-STR Matches

Men who have Y-STR tested with Family Tree DNA will be able to log on to their FTDNA account to see lists of men whose results match theirs and at what step-wise Genetic Distance (see above for definition). Matches are reported at GDs up to 1 for 12 markers, 2 for 25 markers, 4 for 37 markers, 7 for 67 markers and 10 for 111 markers. Men whose STR haplotypes are very close to modal are likely to have many pages of matches, particularly at Y12, Y25 and Y37. Those with unusual haplotypes, far off-modal, will have few matches. The fewer matches you have listed and the smaller their GD from you the more likely it is that these matches are truly relevant to your genealogical research, being men with whom you have shared paternal line ancestry within recent time. FTDNA provide statistics relating Genetic Distance to the probability of a relationship being within a certain number of generations for men with the same surname. For men with different surnames the probabilities will be less than these.

https://www.familytreedna.com/learn/y-dna-testing/y-str/two-men-share-surname-genetic-distance-37-y-chromosome-str-markers-interpreted/

https://www.familytreedna.com/learn/y-dna-testing/y-str/two-men-share-surname-genetic-distance-67-y-chromosome-str-markers-interpreted/

https://www.familytreedna.com/learn/y-dna-testing/y-str/two-men-share-surname-genetic-distance-111-y-chromosome-str-markers-interpreted/

Note I have not included links to FTDNA's 12 and 25 marker matches statistics. Experienced genetic genealogists do not have any confidence in 12 marker matches. Twelve is simply too few markers to provide reliable matches. Many 12 marker matches will be due to chance convergence of haplotypes of men who have no shared paternal line ancestors within the last several thousand years. Ignore 12 marker matches. The same applies to 25 markers unless you happen to have a large number of infrequently occurring allele counts within your first 25 markers (as in my kit number 170191 results - see above for my haplotype data). For most men the number of matches reduces significantly from 37 to 67 markers and confidence that the remaining matches are

genealogically relevant increases accordingly. Once you have identified likely relevant matches, use the e-mail link on the Matches web-pages to contact your potential genetic cousins explaining why you are contacting them, sharing relevant information about yourself and asking them to provide the same to you. If they prove to be S16264+ men like you, please invite them to view the www.S16264.info web pages and to join our discussion forum.

You may also be able to more clearly identify relevant matches if you make use of the YSearch genetic genealogy database. FTDNA provide a link on your results pages for uploading your results to YSearch. I recommend you do this and then use YSearch's "search for matches" facility. The advantage of using YSearch is that you will be able see your matches haplotype data and some brief notes on their family history which can help confirm that the match is likely to be helpful to your research. YSearch also allows you to search for matches at greater distance than the limits used by FTDNA and to search for men with specified allele counts which you are confident are specific markers for your paternal line.

The absolute proof that a match is genealogically relevant requires more than just a Y37 or higher match: it needs to be supported by other evidence. This additional evidence can include a shared surname, a known relationship from traditional family history research, or additional genetic data such as a shared recent Y-SNP or, less conclusively, an autosomal DNA match.