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

S16264 and Y-DNA SNP Tests

Prepared by Nigel Bond

Version of 10 June 2018 – Check here for updates

To confirm that you belong to R-S16264 at least one of the males in your paternal family will have had a detailed test of his Y-DNA SNPs. That test will have indicated that he is S16264+ and/or S21225+, either by direct testing of those SNPs, or indirectly through a positive result for one of S16264's descendant SNPs (e.g. L679). Alternatively, you may have taken an STR test such as FTDNA Y37 or Y67 that shows you share recent close paternal ancestry with men who are S16264+.

If you would like to find out more about Y-DNA fundamentals and Y37 and similar tests, please refer to the Background Information pages.

Once you have confirmed that you are S16264+ and/or S21225+, you might choose to carry out further SNP testing to find more recent SNPs that you share with others within your family name group and R-S16264. These data will allow us to understand more precisely how the different family groups within R-S16264 are related to each other, and will help to estimate when their most recent common ancestor lived.

Once we have this information from a number of R-S16264 families we will be able to develop a detailed family tree and history of our R-S16264 clan. Ideally every family group will have at least one man tested to determine his most recent SNPs.

Commercial Y-DNA SNP Tests

There are four main types of Y-DNA SNP tests available for genealogy:

1. **Phylogenetic tests** such as **National Geographic's Geno 2.0**. These determine a man's position on the Y-DNA phylogenetic tree by testing thousands of known SNPs in a single test. These tests typically cost of the order of **US $200-300**. Note though that they do not include **all** known SNPs. Instead, they are largely focused on the more ancient SNPs together with a few frequently occurring more recent SNPs. Geno 2.0 does not include S16264 or S21225. Phylogenetic tests of this type are not recommended for men who already know they are, or almost certainly are, S16264+.

2. **Individual SNP tests** as offered by **Family Tree DNA** and **YSeq** where the sample is tested for a single named SNP. These cost **US $39** per SNP at FTDNA and **US $18** per SNP at YSeq. Such tests may be of benefit for S16264+ men for answering specific research questions.

3. **SNP panels** covering around 150 known SNPs. These are designed to allow men who know approximately where they belong on the Y-DNA phylogenetic tree to test for a number of more recent SNPs and so determine their specific clade or sub-clade with greater precision. These

panels are offered by **Family Tree DNA** and **YSeq** and typically cost of the order of **US $100-140**. The first version of FTDNA's **M343 Backbone SNP Panel** Test proved to be very effective for identifying new S16264 families. Unfortunately the current version does not include S16264. It now requires two FTDNA panel tests for a M343 man to discover he is S16264: a M343 Backbone Panel followed by their L21 Panel.

4. **Next Generation Sequencing (NGS)** tests such as **Family Tree DNA's Big Y-500** and **Full Genome Corporation's Elite** determine the detailed nucleotide sequence (e.g. -AGATTAC- etc.) of very large segments of a man's Y-DNA. The nucleotide at each position is checked against a standard reference sequence (see Note below) to determine where SNPs have occurred. Where a SNP is found, it is either named if it is already known or it is identified as a new SNP discovered in this NGS test. The NGS test therefore determines the man's position on the known Y-DNA phylogenetic tree and discovers new SNPs not currently on the Y-DNA phylogenetic tree.

Typically a new SNP mutation occurs every 3 to 5 generations so a NGS test of a typical S16264 man will find more than 30 SNPs which are more recent than S16264, some of which will be shared with other S16264 men and some which will be currently unique to that man and his close paternal line male relatives.

NGS tests are the best available for genetic genealogy research but are expensive, currently in the range **US $500-1000 US**. Prices are expected to fall as the technology becomes more efficient. FTDNA also have sales at fairly regular intervals (Holiday Season, Fathers' Day, etc.). In June 2018 men who already have Y111 were able to purchase Big Y-500 for $349.

The S16264 Project's preferred genetic genealogy testing companies are Family Tree DNA, YSeq.net and Full Genomes Corporation because these companies' Y-DNA tests allow results to be compared on a consistent and comprehensive basis.

Interpretation of SNP Test Results

The interpretation of the results of test types (1) to (3) above is reasonably straightforward. The results will be in the form of a list of SNPs with a result for each: either negative (ancestral) or positive (derived). These will enable you, the tested man, to position yourself on the phylogenetic tree with the level of precision corresponding to the most recently occurring positive (derived) SNP tested. There may also be a written report with your testing company's interpretation of the results. If you have questions members of this group can help.

The interpretation of type (4) Next Generation Sequencing Results is much more complex and requires more effort by the tested person. Fortunately there are volunteer experts as well as paid interpretation services who are ready to help. Keep in mind the main reasons for taking an NGS test:

(1) to determine where you and your close paternal line family are placed on the Y-DNA phylogenetic tree including SNPs from ancient times down to the present day,

(2) to understand how closely you are related to other family groups by comparing positions on the phylogenetic tree and to see if that gives clues on shared geographic origins,

(3) to estimate the dates at which branches in the tree occurred by counting SNPs in descendant blocks and individuals and applying an estimated SNP mutation rate,

(4) to use the above to develop a genetic history for your family and for other S16264 men linked to known family history and, if possible, to past events in the historical and archaeological record.

(5) to contribute to genetic genealogy research by adding new branches to the phylogenetic tree

(6) to provide a more detailed reference tree which will enable future testers to determine their relationship to you and vice versa

The most important of the volunteer experts is Alex Williamson (www.ytree.net). Alex analyses NGS results for P312+ men (who include S16264+ men), comparing them to others and then placing them on his phylogenetic tree, which includes the S16264 branch. Your first action on being told that your results are available should be to transfer your results to the Y-DNA Data Warehouse following the instructions given at http://www.haplogroup-r.org/instructions.html. Alex will download your data from the warehouse.

Alex's help is necessary because of the complexity of the tree and because NGS tests do not always give clear results: sometimes SNPs will be detected with 99.9% confidence, sometimes at lower confidence, and sometimes a region of the Y-chromosome will be read clearly in one man's test but not in another. So an expert eye is needed to independently determine whether the differences between two men's results are real or if they are attributable to differences in the effectiveness of the SNP detection for the two tests. You can see the detailed results of Alex's analysis by clicking on one of the blocks (long lists) of SNPs on his tree and then clicking on "view the mutation matrix for these men" to reveal a matrix showing how effectively each SNP was reported for each man: only the SNPs labelled '+' were detected with full confidence.

There are also two paid interpretation services you should consider using: YFull and Full Genomes Corporation (of course FGC's interpretation is included if you have purchased one of their tests). Both cost 50$ for FTDNA L21 Project members. Both extract more information than FTDNA from the FTDNA BigY raw data files, finding more SNPs (three of my most recent SNPs, the SNPs most useful to me for determining relationships within the past 600 years, were found in my BigY results by FGC and not by FTDNA). YFull also provide date estimates for when branches occurred in the tree, presenting that data on their tree. You can download your raw data file (your .bam file) from your FTDNA BigY results web page, similarly to the way you downloaded the .vcf file for Alex. If you use these services you should share the interpretation data with Alex so he can use it to enhance the S16264 tree.

Some people also choose to examine their detailed .bam files themselves using specialist software which can be downloaded from the Internet. This is a highly technical exercise and is beyond the scope of this guidance note. It is easier to review the .vcf file: it can be opened as a spreadsheet using Excel. Studying the .vcf file gives you the opportunity to explore and better understand your results.

The above guidance applies whether you have purchased a FGC NGS test or FTDNA's 'Big Y-500'. FTDNA customers also have the opportunity to view their results by logging in to their FTDNA account. When, in 2017, FTDNA changed the Y-DNA reference sequence to Build 38 they took the opportunity to upgrade their on-line Big Y reports which had previously been very poor. Features now include a chromosome browser where the SNP reads are displayed graphically. SNP results can also be seen in two tables: Named Variants and Unnamed Variants. '*Named*

Variants' displays your SNPs that are on the list of ~70,000 known SNPs against which Big Y data is compared. '*Unnamed Variants*' displays your SNP markers that are *not* on the list of ~70,000 known SNPs. These markers may or may not be unique to you as an individual. Men in related lineages may share some Unnamed Variants. As men from distantly related lineages test, SNP markers may be moved from Unnamed Variants to Named Variants. There is also a '*Matching*' tab which displays your terminal SNP matches. Your terminal SNP determines the terminal (final) subbranch (on the Y-DNA haplotree) to which you belong. A person is considered a match if they have 30 or fewer differences in SNPs with you, and their haplogroup is downstream from your haplogroup or downstream from your four closest parent haplogroups.

Developing a family phylogenetic tree

Once several men in a family group have NGS tested it is possible to ask YSeq.net to develop Sanger tests for a selection of the family's recent SNPs. Other family members may then purchase individual SNP tests from YSeq to determine their relationship to the NGS tested men. The family may then develop a family phylogenetic tree using SNP and STR haplotype data without the need to incur the cost of additional NGS testing.

Notes:

1. Prior to 2017 all the testing companies were using a version of the reference sequence called Build 19 (aka Hg19). The master reference sequence used by human genome researchers is continuously updated to reflect latest research findings. Over time Build 19 became obsolete and, in 2017, was replaced by Build 38 (aka Hg38). This has not resulted in any significant changes to the Y-chromosome phylogenetic tree, however some DNA segments assigned to the Y-chromosome in Build 19 have been found to belong to the X-chromosome and, to a lesser extent, vice versa. As a result the position numbers for most Y-SNPs changed: for example, S16264 moved from position 14,658,766 in Build 19 to position 12,546,831 in Build 38. It is therefore important to understand which Build is being used as reference when quoting SNP positions or when SNPs are labelled as Position-Ancestral-Derived.

2. As well as SNP data, NGS tests report a large number of STR results. For example, FTDNA's Big Y-500 reports up to 450 STRs not included in FTDNA's Y37, Y67 or Y111 tests. However most of these 450 STRs are unlikely to add much useful information for phylogenetic studies as they appear to be very stable with few mutations.