R1b-S16264: Information for Newcomers

This information is divided into two parts:

STRs, SNPs and Phylogenetic Trees (this page)
and
DNA STR Tests (click here or follow the link below)

STRs, SNPs and Phylogenetic Trees

Prepared by Nigel Bond

Version of 10 June 2018 – Check here for updates

Genetic information is stored within our **DNA** in a code made up of sequences of four **amino-acids**. These amino-acids (also known as DNA **nucleotides**) are identified by the initial letters of their names: **A**denine**, G**uanine**, C**ytocine and **T**hymine.   A strand of DNA might, for example, contain a sequence such as -AGTACGG-.   Our DNA is stored in biological entities called **chromosomes**. The **Y-chromosome** occurs only in males and is passed, usually unchanged, from a father to his sons. Occasionally errors (**mutations**) occur in the transmission from father to son that are of no significance biologically or medically but which will then be passed on to all future descendant generations. Y-chromosome genetic genealogy is based on tracking two types of these benign mutations: **STRs** and **SNPs** (pronounced 'snips').

STR stands for **Short Tandem Repeat** in which short sections of the Y-chromosome genetic code are repeated multiple times. For example, a father might have a sequence, such as -AGAT-, which is repeated 10 times (his '**allele count**' is 10) while his son has 11 repeats (his allele count is 11) of the same sequence at the same location on his Y chromosome. All the son's descendants will have the same 11 repeats until another STR mutation occurs at this location. These STRs occur at specific locations in the Y-DNA known as '**markers**', each identified by a specific **DYS** number such as DYS455. STR DNA testing counts the number of these repeats at a set of marker locations. For example Family Tree DNA's Y37 tests 37 markers, Y67 tests these 37 plus 30 more, Y111 tests the 67 plus 44 more. The results are reported as a series of 37, 67 or 111 numbers (the allele counts) in a standard sequence. This series of numbers is the tested person's Y-DNA **STR Haplotype**. For example, a 12 marker haplotype might be '13 24 14 11 11-14 12 12 12 13 13 29'.

SNP stands for **Single Nucleotide Polymorphism** in which a single amino-acid in the DNA is replaced by another, for example a Cytocine by a Thymine. SNPs are identified by a letter and number combination where the letter identifies the laboratory which originally discovered the SNP. For example, the S in our S16264 indicates that this SNP was discovered by Dr. Jim Wilson's laboratory at BritainsDNA. A SNP can also be identified by its location on the Y chromosome, the **ancestral** (original) nucleotide and the **derived** (after mutation) nucleotide. For example SNP S16264 is at location 12,546,831 on the Y-chromosome with ancestral nucleotide Adenine and derived nucleotide Guanine and so can be written as 12546831-A-G or 12546831-A>G. Men who carry the derived nucleotide of a SNP are **positive** for that SNP (for example, all men qualified to

join our group are S16264+ and/or S21225+) while those who carry the ancestral nucleotide are **negative** for the SNP (men who are not qualified to join our group are S16264- and S21225-).

Most SNP mutations have occurred only once in the history of mankind. Such a SNP is known as a **Unique Event Polymorphism** or UEP SNP. Because these UEP SNPs are inherited by all men who are descendants of the first carrier of each SNP (their **Common Ancestor**), successive UEP SNPs can be used to draw the family tree of mankind with each branch defined by its initial UEP SNP. This tree is known as the Y-DNA **Phylogenetic Tree**, or **haplotree**. The ancient branches on this tree and their descendant branches form the Y-DNA haplogroups. S16264 defines a relatively recent sub-branch within the ancient Haplogroup R. **Haplogroup R** is defined by SNP M207 with the first carrier of this SNP living more than 24,000 years before present (YBP): this being the radiocarbon date of the remains of one of his descendants, a M207+ boy found at Ma'lta, near Lake Baikal, Siberia. One of M207's immediate descendants, R1, is defined by SNP M173+, which in turn has descendant haplogroup R1b, defined by SNP M343+. The line of descent leading to S16264 is:

R-M207+ > M173+ > **M343+** > L278+ > L754+ >L388+ > P297+ > **M269+** > L23+ > L51+ > L151+ > **P312+** > Z209 > **L21+** > **DF13+** > Z39589+ > **S16264+**

Estimated ages for the most recent branches in this line are P312 5000-4700 YBP, L21 4800-4500 YBP, DF13 4700-4200 YBP, S16264 4200-3800 YBP. These are very approximate estimates and are not agreed by all researchers. The ages can be expected to change as research progresses. However we can say with reasonable confidence that **the first S16264+ man was born in the European Bronze Age, approximately 2000 BCE (+/- 500)**.

See the **International Society of Genetic Genealogists** (ISOGG) web-site http://www.isogg.org/tree/ISOGG_HapgrpR.html for the fully detailed, current and most widely referenced version of the haplotree. You will see on this tree that there are often several SNPs listed at each branch of the tree separated by commas: these SNPs are **phylogenetically equivalent**, which means that we do not know the sequence in which they occurred (we can only start to place SNPs in sequence when we discover a man who is positive for one or more SNPs and negative for the others, proving that his positive SNPs came first – see note 5 below). **S21225** is phylogenetically equivalent to S16264. You will also see SNPs with several names (separated by a /) as a result of more than one laboratory claiming to have discovered the SNP independently. For example L21 is also known as S145. The groups of men descended from more recent branches of the Tree are termed **clades** and **subclades**. S16264's subclade is labelled **R1b1a1a2a1a2c1a1g** on the March 2017 version of the ISOGG Haplotree, with each letter and number corresponding to a branch in the tree.

Notes

(1) Unfortunately testing companies' haplotrees are generally not identical to the ISOGG Tree, particularly for relatively recent branches. Also, testing companies often use their own names for SNPs in preference to more widely recognised names which can make comparing trees unnecessarily difficult.

(2) The clade and subclade labelling system used by ISOGG with alternating letter-number-letter-number for each branch creates some confusion as every time a new intermediate branch is discovered the labels change on all descendant clades and subclades. For this reason researchers are increasingly using SNP names instead of the ISOGG system. So R1b1a1a2a1a2c1a1g becomes **R-S16264**. This also overcomes the issue of differences between different haplotrees.

(3) SNPs which are known to have occurred independently at different locations in the phylogenetic tree are known as **recurrent** SNPs.

(4) Another type of mutation which may be useful phylogenetically is the **InDel**, or **Insertion-Deletion** in which a nucleotide or string of nucleotides is inserted in (or deleted from) the Y-DNA.

(5) In February 2017 a man was tested whose results split L21's phylogenetically equivalent block in two – a new upstream (ancestral) block headed by Z290/S461 with three equivalent SNPs and a descendant block headed by L21/S145 with two equivalent SNPs. The ISOGG Haplotree has been revised to reflect this.

(6) Until 2017 S16264's reported position on the Y-chromosome was 14,658,766; some web-sites will still describe S16264 as 14658766-A-G. In 2017 the reference sequence for the Y-chromosome was updated to reflect current research and this shifted the position of S16264 from 14,658,766 to 12,546,831 and so its alternative description became 12546831-A-G.

Onward to Part II: Y-DNA STR Tests