

Home

Messages


# Hashtags

R1b-U106@groups.io / Topics / SNP tracker

Search

## SNP tracker

Date 1 - 7 of 7

 iain

1/19/21 #3805

Hi James, all,


Rob's SNP tracker takes its SNP ages directly from YFull, so those ages are nothing to do with me.

Geographically, the SNP tracker is another question. I've been in fairly close touch with Rob about these methods. The problem of geographical spread is a hard nut to crack, and no-one's done a good job of this yet. Generally speaking, the geography of individual groups in the SNP tracker is too far west, but there are all sorts of biases and uncertainties in the data that aren't easily dealt with. I might have a go at this theory once FTDNA and I are done with the ages.

Cheers,

iain.

More

 Mark Miller

1/19/21 #3806

iain, Mark Miller here. New to this forum. When do anticipate ages project with FTDNA will be completed?

Show quoted text

More

 iain

1/19/21 #3808

Hi Mark,

Welcome!

At the moment, the only information that is public knowledge is that FTDNA are working on adding ages of haplogroups to their BigY product, and I am helping them with the background mathematics and statistics as part of that.

I am privy to certain information as a result of that help, but which I can't discuss on this forum. However, that additional knowledge doesn't include a clear timeline of when any release is going to take place.

The process is taking a little longer than I had anticipated, but that's often the case in research and development. What I have seen so far is very promising and, while they haven't taken all my recommendations on board for various reasons, I am pleased to see they are doing a rigorous job of making sure this works, and squashing any problems I have thought of.

History as taught me that no product ever works perfectly for everyone, but I have good reason to believe that the dates they're going to provide are going to be both more accurate and precise than a haplotree-wide set of dates I am able to come up with on my own, because they have done a good job of checking everything works, and because they can run this on their entire database. So I am happy with their progress so far.

Cheers,

iain.

More

 Mark Miller

1/19/21 #3809

That is very encouraging, iain. Thank you.

Show quoted text

More

 Geoff Blackburn

1/19/21 #3810

Thanks again iain,  
This is indeed great news and will be a significant development. I like the idea of "they have done a good job of checking everything works, and because they can run this on their entire database."  
regards  
geoff

Show quoted text

More

James Miller Kit 356313 &lt;millerdriller@...&gt;

1/21/21 #3812

Hello iain,

Hope this finds you and yours well during this time that almost equals the time of the Bubonic Plague Era.


My question would still be can I trust the age estimates of SNP Tracker? I don't know this for sure but have read, somewhere, that your calculations are used in said. This is pure speculation so I would like to hear it from you, yes or no, can I trust SNP Trackers SNP age analysis to be at least close? I do realize that their point of origin for given SNP's may not be altogether exact but that is not my question.

Sincerely,  
James

Sent from Mail for Windows 10

Show quoted text

More

 iain

1/21/21 #3813

Hi James, all,

As I said, the ages in the SNP tracker come from YFull, not me. There is none of my age estimation in the SNP tracker as far as I am aware. I do discuss methodology with Rob, but that's it. You're better off going back to the original data source.

Whether you want to trust the ages that are in YFull is another matter. The basic principle behind their calculations is sufficiently sound - it's the same calculation we all use for SNP-based ages:

$$\text{age [in years]} = \text{number of mutations} * \text{number of base pairs tested} / \text{mutation rate [per base pair per year]}$$

The only things that differ between the different SNP-dating methods are:

- how the number of mutations is counted,
- how to count the number of base pairs tested,
- the exact mutation rate used,
- how to add a layer of statistics on top to account for the uncertainty and biases that come from the randomness of mutations,
- the underlying set of tests that are used, and
- whether any additional data is taken into account.

The first two don't really matter, so long as they're done consistently with each other and in a way that is valid for the mutation rate that's used. YFull are now accepting VCF data from Family Tree DNA, and the definition of the test coverage that goes into the BED file is very low. Consequently, either YFull is probably slightly over-estimating either the test coverage or the number of mutations (by including poor-quality false positives in the calculation). Either should make their ages slightly older than they actually are. Whether or not this is a significant effect, I can't tell.

I have some qualms about how their statistical package accounts for uncertainties. In particular, the documentation suggests that they use Gaussian errors, rather than full Poisson statistics. They also have poor treatment of causality, and have to manually adjust their tree to avoid child clades being older than their parent. There are statistical workarounds for this, but they're neither pretty nor exact. This will lend bias to some branches, and mean that the ages between parent and child branches aren't always trustworthy, even if the overall ages are approximately correct.

YFull only contains a small subset of the overall data that's available to Family Tree DNA. This means their underlying set of tests is small, and their uncertainties are correspondingly large. Potentially, the most serious consequence of this - and I don't know how Rob deals with this - is that haplogroups that are on YFull's tree don't always match up with those on Family Tree DNA's tree, even when they have the same name. This is because many of those haplogroups have been split by FTDNA. I also don't know exactly what Rob does for haplogroups that don't have ages in YFull - I presume he just counts SNPs down the tree, but he'll have to do this without knowledge of whether those SNPs come from BigY-500 or -700 tests, which makes a big difference.

YFull's uncertainties also remain large because they only take SNP data into account. If you take STR data and any other historical information you can get your hands on (paper trails, surnames, ancient DNA), then you can create much more accurate results... at least, in theory.

Most of these differences don't matter for haplogroups further up the tree. So YFull's ages for haplogroups like R-U106 work out as well as anyone else's. These are constrained by archaeology fairly well anyway, and I don't know to what extent YFull's manual edits have fudged their ages in the direction of these archaeological results. The main locations these differences affect are near the tips of the tree, where lack of data and precise statistical uncertainties matter a lot.

However, none of these differences matter anywhere near as much as the random uncertainty generated by the mutations themselves. Remember mutations don't work like clockwork, and that leads to "HUGE" uncertainties, especially when you start looking at historical-era haplogroups. If you are looking just at the ages themselves, without looking at the range of uncertainty that surrounds them, you might as well not be looking at all. Despite the minor grievances with YFull's methods that I've listed here, most of the true ages of their haplogroups will still fall within the range of uncertainty that they've defined - they're just not quite as accurate as they could be.

Cheers,

iain.

More

[← Previous Topic](#)
[→ Next Topic](#)

1 - 7 of 7