

Journal: [www.joqq.info](http://www.joqq.info)

Originally Published: Volume 9, Number 1 (Fall 2021)

Reference Number: 91.003

## WORD TO THE YS: UNDERSTANDING Y-DNA STR GROUP ANALYSIS THROUGH ANALOGY

Author(s): *J. David Vance*

---

## WORD TO THE Ys: UNDERSTANDING Y-DNA STR GROUP ANALYSIS THROUGH ANALOGY

By J. David Vance

“Word to the Ys” is a new feature series where guest authors write about experiences and methods for using data from Y-DNA testing in the practice of genetic genealogy.

### Introduction

I apologize in advance to those who don't like analogies, but I find them useful; and especially powerful when people are struggling with a particular concept and the analogy forces them to think about the concept in a different way that helps them make sense of it. Analogies are never perfect of course, so you have to take them in context of whatever point they were being used to make and know that they are probably not otherwise applicable.

This analogy is really for people who may be still struggling with STR allele values as numbers and trying to understand how they can be used to analyze a subgroup of a surname project (the same techniques can be used with subgroups of haplogroup projects or other types of projects, but it's easier to explain using a surname project).

Please understand that this is only a simple introduction and the analogy is by no means perfect. It certainly offers no insight into the biology of DNA, it glosses over some of the complexities of STR analysis, and it does not address SNPs (like you get from Big Y700 testing or Whole Genome testing) at all.

Also I'm just explaining a way to use STRs, not advocating that using them alone is the best strategy. In fact I usually advocate for using both SNPs and STRs if at all possible. Many people want to jump straight into SNPs and ignore STRs completely, and while there is no doubt that SNPs can give structure and more certainty to a group analysis if you can include them, there are many examples where STR analysis can be useful, like:

- If you've taken on the task of analyzing a group that has already extensively tested STRs, and you want to make best use of the investment they've already made in testing before asking them to upgrade to Big Y700 or other SNP tests;
- If your group doesn't have the available funds to do extensive SNP testing;
- Or if the group has already done extensive SNP testing but key questions still remain about the branching within the group because no SNPs have yet been discovered that mark those branches. STRs can often be used to find additional branching in between SNPs.

But hey, if you'd prefer to ignore STRs completely and only work with SNPs then that's fine and this analogy is not for you. Feel free to skip it.

Before we start out with a group analysis it's also important to know what we're trying to find out. In this analogy we have a surname project subgroup that wants to know how they are all connected to each other; this is a common question but by no means the only one. Are they trying to break through specific brick walls? Are they trying to connect back to a particular ancestor of some importance? Are they trying to show a lineage to a particular ancestral group like a clan or specific community? Are they trying to find a common region of origin? These will all guide your purpose in analyzing the group and what information is important to look at, because it's probably why they invested in Y-DNA testing in the first place. Of course the other possibility is that you paid for their testing for a specific purpose, in which case that's the purpose that's important. So again in this example we'll be trying to build a genetic family tree that connects a group that don't already know how they're connected, but take that as just one example among many and tailor your approach to what your group is interested in finding out.

Ok so with that preamble over, here's the analogy.

### **The Analogy: STRs as words in a Nursery Rhyme.**

Suppose a man long, long ago taught the nursery rhyme "Humpty Dumpty" to his sons, and started a tradition in which every one of his male descendants taught that same nursery rhyme to his own sons in turn. The tradition has never been broken, but occasionally a father mixes up a word here and there in the rhyme when teaching it to one of his sons, so the rhyme can change slightly as it is passed down.

So this first ancient ancestor taught his sons "Humpty Dumpty sat on a wall, Humpty Dumpty had a great fall, all the king's horses and all the king's men couldn't put Humpty together again." And his sons taught their sons the same rhyme, but maybe one of his sons mis-remembered and said "Humpty Dumpty sat on the wall" and so on, and his own sons taught it that way to their sons, and over time the rhyme has been slightly altered on certain male descendant lines from this original man.

As we'll see in a minute, the other important part of the analogy is that some words in the rhyme are more easily changed than others. "Wall" and "Fall", for instance, are easy to remember because they sound the same and because they're important to the meaning of the rhyme, so you wouldn't think a man would change those words very often. On the other hand, a determiner like "a" could easily be changed and you could easily see how instead of "a wall" a man could teach his son "the wall" or even "some wall". In our analogy, words that don't change often will take the place of slow-mutating markers, and words that can more easily change will take the place of fast-mutating markers (and I'll note that often it's not really important what the exact mutation rates are; for analysis purposes it may just be important to see if you're working with particularly fast-mutating or slow-mutating markers).

It's now present day, and you have gathered together 10 of the descendants of that original man, who have all done some level of genealogical research but they don't know how they're all connected. You test them each to see what they remember of the start of the nursery rhyme, and 8 men take a "12 word/marker" test and 2 men take only a "6 word/marker" test (for simplicity, I've used fewer "markers" than actual Y-DNA STR tests, but you get the idea). You put all the results in the

following table. For ease of reference, we'll label these first 12 words and we'll call them "DYS001" through "DYS012" so we can refer to them later.

Kit Number	Paternal Ancestor Name	DYS001	DYS002	DYS003	DYS004	DYS005	DYS006	DYS007	DYS008	DYS009	DYS010	DYS011	DYS012
Group 1													
<b>MODE</b>		Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900001	John Ancestor, b. 1750	Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900002	William Othername, b. 1910	Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900003	Unknown (adopted)	Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900004	Matthias Ancestor, b. 1695	Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900005	Thomas Ancestor, b. 1805	Humpty	Dumpty	sat	on	a	wall	Humpty	Dumpty	had	a	great	fall
900006	Matthias Ancestor, b. 1695	Humpty	Dumpty	sat	on	the	wall	Humpty	Dumpty	had	a	large	fall
900007	William Ancestor, b. 1892	Humpty	Dumpty	sat	on	the	fence	Humpty	Dumpty	had	a	large	fall
900008	George Ancestor, b. 1855	Humpty	Dumpty	sat	on	the	fence	Humpty	Dumpty	had	a	large	fall
900009	George Ancestor, b. 1855	Humpty	Dumpty	sat	on	a	fence	Humpty	Dumpty	had	a	large	fall
900010	Henry Ancestor, b. 1721	Humpty	Dumpty	built	a	great	wall	Humpty	Dumpty	had	a	huge	fall

Figure 1. The group of descendants with up to 12 markers tested.

(and yes, this picture is formatted on purpose like a project's DNA Results page just to make the analogy clearer that we're using words in place of the STR numbers that you'd usually see).

From a regular genealogy standpoint, two of the men have a different surname, but one of the two was adopted anyway and doesn't know his biological father. The rest of them all carry the same surname ("Ancestor"), and several have traced their ancestry back to the same earliest ancestors but those are all in different centuries and while it's clear that they're probably all related, we don't yet know how.

In our picture above, we only see the earliest known paternal ancestors for each of the kits, but as a project administrator it's important also to collect the whole known paternal lineage from each member including the regions where their ancestors lived and any information, even hints, that they might have about their male paternal line. I'll give a few examples in the rest of this introduction about

how this information can help us in building our genetic family tree.

So now let's stare at the "markers" for a bit.

The "Mode" row, also called the "modal haplotype" or just "modal", is the collection of the most commonly recurring words among the group, and clearly "Humpty Dumpty sat on a wall, Humpty Dumpty had a great fall" is the "modal haplotype" for this group. Because words don't change very often, the most common set of words is likely to be the set of words that the common ancestor of this group first taught his sons, which is known as the "ancestral haplotype". It's not guaranteed that the modal haplotype and the ancestral haplotype are always the same, but it's usually a good starting assumption especially if you have more than a few testers, and normally if the modal and ancestral are different it's only on one or two markers. In this introduction because we're keeping it simple, the Mode will be the same as what the common ancestor taught his sons. Of course real STRs also

have a Min and a Max row which can occasionally come in handy, but since words don't have a "min" or "max" I've left those out here.

So then we have "off-modals", meaning words that are different from the Mode values. These are highlighted in red and blue in the earlier picture. Since STRs are numbers of course those can go up and down; in this analogy we just have words that have "mutated" to other words.

So how do we put this puzzle together? Can we tell how these men are related to each other?

## Looking for Strong STR Patterns

Let's start with kits 90007, 90008, and 90009, because we can immediately see that they have a lot of "off-modals" in common. They all have "fence" as an off-modal for marker DYS006, and the two who have tested out to 12 markers have "large" for DYS011 instead of the modal "great". Changing "wall" to "fence" is not an easy mistake and it's not likely to happen very often (an analogy for a very slow-mutating STR marker), so it's very likely that only one father some time in the past said "fence", and his sons passed that mutation of the rhyme on to their own descendants.

It's very tempting to throw 90006 into this analysis immediately also because he shares the off-modals "the" for DYS005 and "large" for DYS011 with kits 90007 and 90008. But he doesn't have "fence" for DYS006, and 90006 also apparently shares the earlier Mathias Ancestor b. 1695 as an ancestor along with kit 90004, who doesn't have "the" for DYS005. So DYS006 is likely closely related to this 90007/90008/90009 subgroup, but let's look at the smaller subgroup of those three first and then we'll widen our search (this is a common strategy by the

way; look at small groups that share key markers first and then expand the analysis out to include other kits from there).

We don't know for sure that kit 90009 also has "large" for DYS011 since he hasn't tested it, but after some study we can conclude that it's very likely that he does, both because everyone else who has the off-modal "fence" for DYS006 also has the off-modal "large" for DYS011 and also because kit 90009 has traced his ancestry back to George Ancestor b. 1855 just like kit 90008. This is one example of where it's important for you to collect the known genealogy information for the group so that you can use it to guide your conclusions. For this analogy of course we don't know for sure that their traditional genealogy research is correct, but let's assume you've checked and decided it's reliable and you can safely conclude that the chances are very high that kit 90009 also has "large" for DYS011 at least until you upgrade him to more markers and confirm it.

So kits 90007, 90008, and 90009 all share the "off-modals" "fence" for DYS006 and "large" for DYS011. Both are important words to the meaning of this nursery rhyme and not easy mistakes to make (our analogy for slower-mutating markers) and so a father changing either on its own would be unusual, but the combination of both in this subgroup of 3 kits makes it statistically near-certain that they share a more recent common ancestor with each other than with the rest of this group. When two or more slower-mutating markers like this point to a subgroup of a larger group, this is a recognizable pattern that says these kits are descended together from one or more common ancestors who (perhaps in one man or perhaps over a series of generations) developed this pattern. The stronger this pattern (the more markers that form it, and the slower-mutating they are), the more certainty this provides about the branching.

In many larger haplogroups, experienced project administrators have created what are called “allele frequency tables” which can tell you how rare a particular STR value is within that haplogroup. These can be useful to assess the strength of a particular STR pattern, because rare marker values are especially more likely to indicate common branches. So for example, kits 90007, 90008, and 90009 all have the value “fence” for marker DYS006. That word destroys the rhyming scheme of our “Humpty Dumpty” rhyme since it doesn’t rhyme with “fall”, and so it would be particularly unlikely for a man to teach the rhyme to his son with “fence” for DYS006. With our analogy we could say that “fence” is an example of a rare marker value and so a pattern that includes it, especially in combination with other unusual markers, could be considered particularly strong.

Patterns like this are sometimes called a “STR signature” that marks a subgroup, or just a “unique STR pattern”, or especially for European or British Isles English speakers, it is also sometimes called a “STR motif”. The more markers there are in the pattern and the slower their mutation rate, the more likely it is that they mark a subgroup that sits on its own branch, and very strong “signatures” are just as good as SNPs for that purpose.

Now we look at marker DYS005 and notice that both 90007 and 90008 have the off-modal “the” instead of the modal “a”. Marker DYS005 though is a determiner in the sentence and not a very important word (our analogy for a faster-mutating marker) and certainly it would not be hard for a father to mis-remember that word so we could expect it would easily change – on its own it’s not even close to a strong pattern, but since we know the genealogy for both 90007 and 90008 is reliable and they work back to a common ancestor we can conclude that one ancestor in their common line

(either George Ancestor or one of his immediate patriline ancestors) had the mutation “the” from “a”. If we diagrammed this situation, it might look something like this:

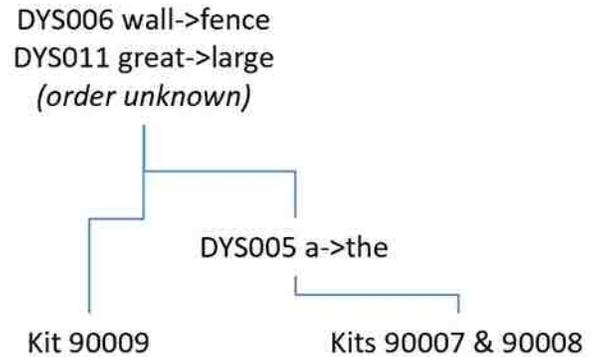


Figure 2. A first mutation history diagram

In the diagram the mutations are marked where they occurred in the ancestral lines, so first one of the common ancestors of all three kits changed “wall” to “fence” and “great” to “large”, and then somewhere after kit 90009 split off, one of the ancestors of kits 90007 and 90008 changed “a” to “the”.

Notice by the way that we don’t know which of the DYS006 and DYS011 mutations happened first; if the same ancestor changed the words for both DYS006 and DYS011 or if they were changed by different men, we only know they both happened somewhere in that range of ancestors. If we did have two or more descendants tested from every generation in this male ancestry then we could easily tell the order of these mutations, but that’s not a realistic expectation. In practice we can only assign certain mutations to certain ranges of ancestors, and the order in which those mutations occurred is unknown (and exactly the same thing occurs with SNPs by the way which leads to equivalent blocks of SNPs).

**Next: Finding the most likely scenario among several possibilities**

Now let's look next at kit 90006, because he shares the "large" mutation for DYS011 and the "the" mutation for DYS005, but NOT the "fence" mutation for DYS006. So he doesn't immediately fit in the branches of our picture above. In fact when we try to include 90006 in our diagram there are three main possibilities for how his version of the rhyme might have come about:

1. Maybe he doesn't fit anywhere in these branches and he's a descendant of a separate line who happened to also make the same two mistakes and also taught their children "large" for DYS011 and "the" for DYS005. This is possible, but statistically not very likely because it means both DYS011 and DYS005 had to have two separate parallel mutations on different branches.

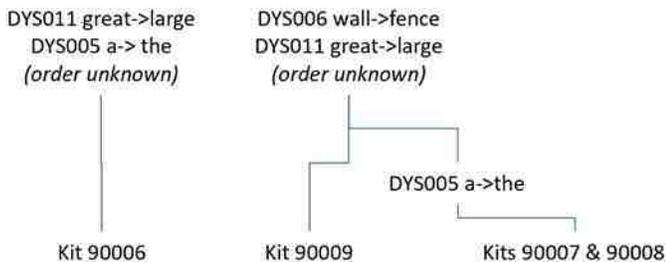


Figure 3. Possibility 1

2. Another possibility is that he is a descendant of the same common ancestors, but one of his ancestors changed DYS006 back to "wall". If that happened, the most likely scenario is that he inherited the "the" word change on DYS005 from a common ancestor

with kits 90007 and 90008, and his DYS006 changed to "wall" from there.

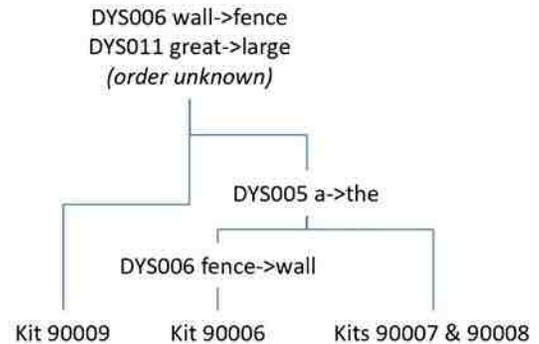


Figure 4. Possibility 2

3. The third and final possibility is that 90006 is related to this group still but earlier before the man who had the "fence" mutation on DYS006. That would mean we've guessed wrong on the DYS005 mutation; that it changed from "a" to "the" first earlier, and it changed back to "a" on kit 90009's line. That still leaves us three kits with "the" for DYS005 and one with "a", but it changes where DYS005 changed among these branches.

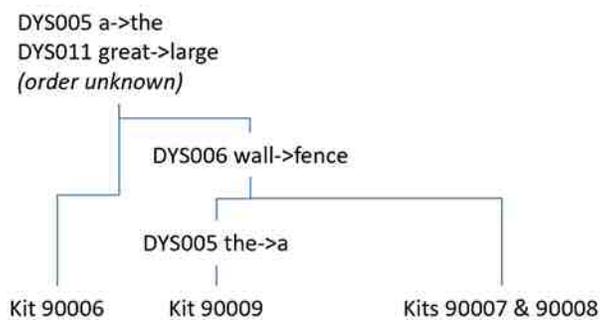


Figure 5. Possibility 3

In our first look at just kits 90007, 90008, and 90009, we found an example of a STR "signature", where two STRs were clearly pointing to a subgroup of men who were more closely related. Now we've moved

beyond “clearly” into probabilities – we have several options that are each possible and we have to narrow them down to the most likely.

This is where many people get discouraged with STRs because they often only give you likely scenarios, not always certain ones. SNPs also actually have a range of likelihoods and don’t always give you complete certainty, but on average their reliability is higher in supporting genealogy scenarios. If I were to diagram the relative reliability ranges of our various sources of information including traditional genealogy research, they might look something like this:

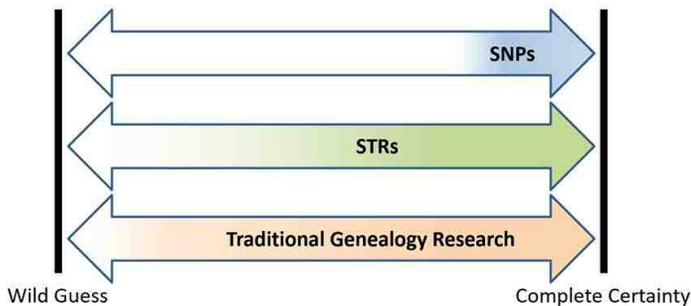


Figure 6. Subjective range of source reliability

With STRs, you can get as close to certainty as SNPs if you find really reliable patterns, but you do have a wider range of reliability in supporting various scenarios that you have to assess. This is not really different from traditional genealogy research, where we’ve always had primary and secondary sources that differ in reliability and have often found that some evidence (like the recollections of the past generations for example) is not always as reliable as we would have liked. So STRs and traditional genealogy research are very similar in that we have to assess them both for reliability, more so than with SNPs.

As I said earlier, perhaps the best approach is to combine all of these sources to answer your group’s

questions. In this introduction I’m focusing on STRs and talking a little about integrating information from traditional genealogy research, but if you do have SNP information from multiple SNP tests (like Big Y700 or Whole Genome testing) among the group this is even better, because SNPs may be able to give you some key sub-divisions that already break the group into clear subgroups even before you start analyzing the STR patterns. Depending on how many SNPs you can find that separate a group, you often can get the high-level skeleton of branches within the group from SNPs and then use STRs supplemented by traditional genealogy knowledge to fill in more detailed branching.

So turning back to our scenario, what can we use to distinguish between these three possibilities?

Well first of all, we notice that each possibility involves markers mutating more than once. In Case #1, DYS011 and DYS005 both had to change twice for a total of 4 mutations; once each on 90006’s separate branch, and once each in the branches that 90007, 90008, and 90009 have in common. In Case #2, only DYS006 had to change twice, and in Case #3, only DYS005 had to change twice. The other thing that we need to remember is that we already said that DYS006 and DYS011 were slow-mutating markers, and DYS005 was a faster marker (using our analogy of the words that are less likely to be changed as being “slower-mutating” than words that could be changed more easily).

Statistically speaking, the scenario that involves the fewest number of mutations is most likely, and markers that mutate at faster rates are more likely to change than markers that mutate at slower rates. Following that logic, Case #1 is the least likely because it results in four extra mutations rather than two. And since we said DYS006 was a slow marker and DYS005 was a faster one, Case #3 is

therefore statistically most likely, because only the faster marker DYS005 has multiple mutations.

When phylogeneticists (scientists who build evolutionary trees) build trees you will often see them applying a principle called “maximum parsimony”, which essentially means trying to minimize the number of changes across the tree. If changes are already unlikely to begin with, then it stands to reason that the fewest number of changes (and additionally, the fewest number of less-likely changes) will result in the statistically most likely tree. But it’s important to also remember that human ancestry doesn’t always follow the statistically most likely path, so if this is the only evidence that we have to go on we can only conclude so far that Case #3 is suggested, but not that it’s true.

The traditional genealogy knowledge might help strengthen either Case #2 or Case #3 also. Kit 90006 knows his genealogy much farther back than any of kits 90007, 90008, and 90009 (and let’s say their earliest known ancestors aren’t in 90006’s line, otherwise they would also know they descended from Matthias Ancestor). That might argue for Case #3 also since 90006’s connection to the others is much further back, but we don’t know enough about that from this example. Perhaps also the regions these ancestors originated from or when their lines immigrated to other countries, etc, would make one of Case #2 or Case #3 either impossible or less likely.

It is also hard to know how much weight to put on the knowledge that both 90004 and 90006 list their earliest ancestor as Mathias Ancestor b. 1695. A lot depends on how reliable the genealogy research is for those two descendants, but it also matters how Mathias may fit into the genealogies of the other earliest known ancestors of this group. Could

Mathias be a direct paternal ancestor of William Ancestor b. 1892 or George Ancestor b. 1855? Or does the traditional genealogy knowledge rule either or both of those out completely? Traditional genealogy research of course as we showed earlier also has a wide range of accuracy from “Wild Guess” to “Complete Certainty”, but the details learned from traditional genealogy research may completely rule out one or more of these scenarios or they may only make one or more scenarios more or less likely. As the analyst of this group it’s up to you to weigh the traditional genealogy evidence and decide that as well.

So perhaps we can say that Case #3 is only statistically more likely, or perhaps more evidence from traditional genealogy agrees or refutes that. For the sake of this example, let’s go with Case #3.

### **Next: Kits that we can’t yet place**

Kit 90004 also traces his ancestry back to Matthias Ancestor b. 1695 like kit 90006, so it’s tempting to add him into our Case #3 tree on the same branch as 90006. Notice that he has “a” for DYS005, so we might have to assume yet another mutation of DYS005 somewhere on our tree.

If we only made a decision based on the rhymes however, there’s no visible reason to put 90004 and 90006 together. It would really be good to know the status of DYS011 for kit 90004 and if he has the signature off-modal “large” for that word, but he hasn’t tested that far. There is no reason off-hand to believe his traditional genealogy research was wrong, but Y-DNA isn’t giving us any real clues either way. If 90004’s traditional genealogy knowledge was rock-solid and we also knew something about Mathias Ancestor’s relationship to William Ancestor and George Ancestor then maybe we could draw

some conclusions from there about how 90004 fits with 90006, 90007, 90008, and 90009 and maybe we could also place 90004 on our little diagram from Case #3 at least as a speculative or even possible branch. For the time being however, let's say we don't have enough reliable information to draw that conclusion.

When we look at kits 90001, 90002, 90003, and 90005, we see we know even less about them. These kits all have the exact modal values for all 12 markers – they received the rhymes unchanged from the group's common ancestor. That might mean that they all belong together on one branch, but it also might mean they belong on different branches that all had no mutations – we don't know. One thing we do know though is that they're not closely related to kits 90006, 90007, 90008, and 90009, since they don't share any of the same "off-modals."

### **Next: The out-of-place kit**

Kit 90010 has a number of "off-modals" that no-one else shares. His ancestors taught him "Humpty Dumpty built a great wall, Humpty Dumpty had a huge fall." All four of those word variations in the rhyme are unique to his line.

In theory it's remotely possible that his ancestry does share one or two of those word variations with

other kits in this group, but after 90010 split off their line "back-mutated" on those words back to the modal values. Without any reliable traditional genealogy knowledge that suggests that though, it's not statistically likely. Leaving our analogy for a minute, with STR numbers of course it's possible that mutations occur in steps in the same direction, so if the modal value is 12 and there is a sub-group that has 13 and another single kit has 14, you may have to consider that the kit that has 14 is a further mutation off the group that has 13. With words we can't illustrate that so we'll leave it out of this example.

Without any other evidence then, all we can conclude is that kit 90010 is most likely descended from an earlier branch of this group.

### **Putting it all together**

So we have a combination of reliable marker patterns and possible scenarios, and if we threw it all together into one diagram it would look something like this:

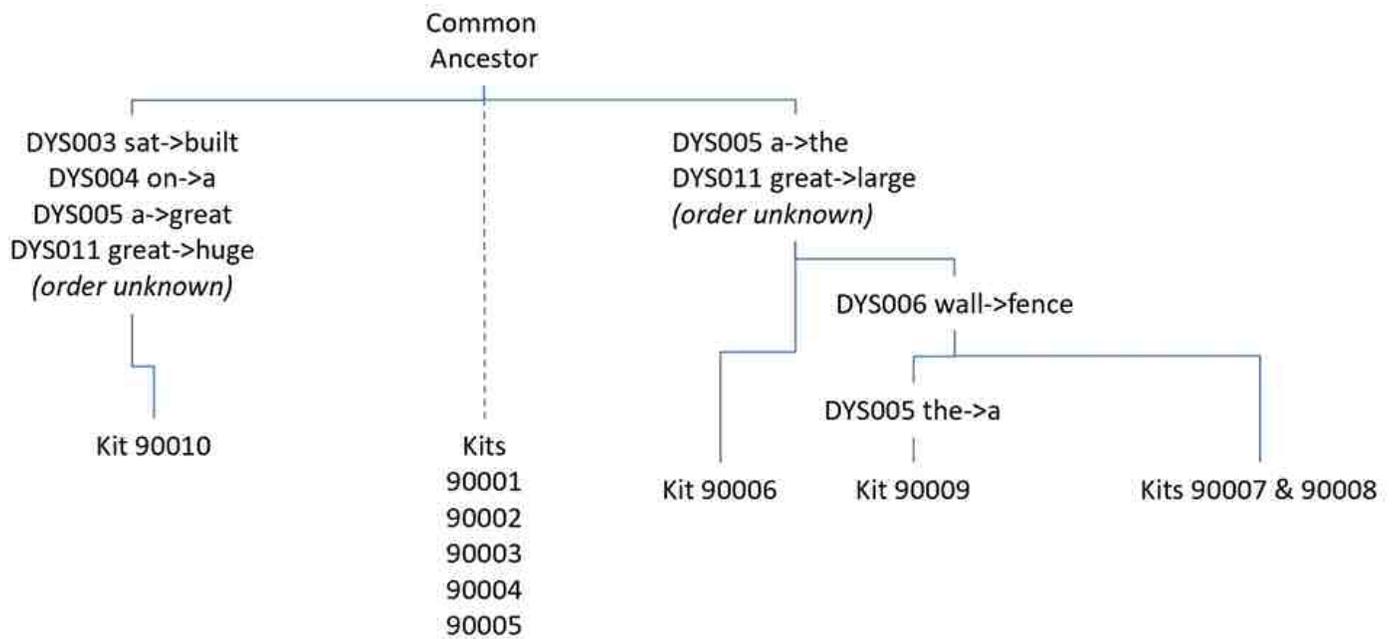


Figure 7. Starting to map the mutations on a tree

If kit 90004's traditional genealogy really was rock-solid of course then we might be able to list him over with 90006 and make an educated guess about how their lineage connected with the ancestry of kits 90007, 90008, and 90009. Since earlier we said we weren't sure about 90004's ancestry we've left that out of this "final" picture, but it's one variation that we might adjust depending on how strong we think that educated guess would be.

This type of tree diagram along with the associated mutation history is often referred to as a "mutation history tree" or "genetic family tree", and at the end of this introduction I'll recommend some resources for additional information and tips.

We can also of course start to add ancestors on the various branches based on what we know from traditional genealogy.

Unfortunately in this example we haven't learned anything about the likely NPE 90002 or the adoptee

90003. Both are probably related to this Ancestor surname group, but this level of information has not given us any clues as to where.

### Next Steps

So where do we go from here? Obviously with this fairly simple analogy and small group of testers, we haven't discovered very much yet about their shared ancestry. But the example is still not unusual in that often you can draw some firm conclusions about a subgroup, that you have possible connections between others in the group, and that for some subset of the group, the DNA hasn't yet helped at all.

We haven't really spent much time in this introduction either on what additional traditional genealogical knowledge we might be able to gather and how it might help us further. Does any of that help guide the branching in more detail? Does it

help us assign ancestor names to the branching points in our genetic family tree? We have to be careful of course if the group is trying to use DNA to confirm their genealogy not to then use the genealogy information they were trying to confirm in the first place, but if their traditional research is reliable enough it may help us in our conclusions about the group's connections.

Adding data is always a good way to learn more. You can target certain upgrades, like in this example upgrading kit 90004 to at least 12 markers would help confirm whether he had the off-modal "large" for DYS011 which might help place him within the 90006, 90007, 90008, and 90009 subgroup. Or if there were higher-level tests that provided more words of the rhyme and you could get a number of these men to upgrade, that might further refine your branching knowledge.

Adding SNPs to the analysis is another obvious improvement, since SNPs can give you the basic branching structure and can quite possibly reliably break this group up into several subgroups. But for instance if you had no branching SNPs that showed you that kits 90006 through 90009 were a separate subgroup, you've at least got the STRs that help you to see that, so combining the two types of mutations is often the best way to get as much from Y-DNA analysis as possible.

You could try using tools to estimate the ages back to the common ancestors for this group – various tools can do this, including (among others) Family Tree DNA's TiP tool, the McGee utility, and SAPP. Age estimation is a complex subject but both STR and SNP-based estimation methods exist and both can only give you very general estimates for the timeframe in which the common ancestor lived, normally with an error range of at least 100 years on

either side of the estimate if it's within genealogical times.

The SAPP tool also provides some automation for creating these genetic family trees even adding in SNPs and genealogy information, although it tries to build the most likely tree from the data and can't show you all the relative likelihoods of alternate scenarios, so if you do use it please consider it as a modeling tool and not a proven ancestry generator. You still have to assess the finished product yourself and decide which branches are more or less likely.

## Further Resources

There are many aspects of STR analysis that I haven't covered in this simple introduction, like methods of more closely estimating the ancestral haplotype, recognizing and dealing with convergence, or adding the extra Panel 6 and 7 STRs found in Big Y700 to the analysis. But if you're just getting started with STR analysis I would suggest you leave those until you've had more practice with STRs in your project.

For more on building genetic family trees using STR mutations, the best source I can suggest is Maurice Gleeson's YouTube videos on the subject, which you should find from a simple search on YouTube using his name and "mutation history trees" (Maurice now refers to them as "genetic family trees," so you may find recent videos under that search term also). Maurice also runs the Genetic Genealogy Ireland channel on YouTube so look for some on that channel also. John Cleary also has many great videos that cover STR analysis as well so his would be another invaluable resource for further study.