# *DNA for your ONS:* Approaches and Expectations

by Susan C Meates MCG (DNA Advisor, Member 3710)

## Our One-Name Studies

Each of our one-name studies is different. Size is the first element that is different, ranging from a very rare surname with a small population to high-frequency surnames. Some surnames are single-origin, though most are probably multiple-origin. Where the surname is found today and the frequency in each country probably varies. Some variants may have thrived, while others may no longer be present in living persons.

Within our one-name studies, the various family trees are also usually different. Some ramified significantly, with lots of males; while others had limited males, and some daughtered out. Migration may have impacted some trees and not others.

Each surname is different, and from the origins of the surname through history to today, many different journeys occurred. For example, perhaps a male from a tree immigrated quite early to the colonies and had many sons, resulting in a large and robust tree in the USA.

Having a one-name study gives you a tremendous advantage when you have a DNA project, as compared to other project administrators who don't have an ONS. In many cases, you have built up information about the surname and the various family trees. This information is invaluable to your recruiting efforts, as well as in analyzing the results from those that have taken a Y-DNA test.

## Expectations

One simple yet informative view of a DNA project is to look at the number of participants as compared to the population of the surname.

If for your ONS there are 360 males in the world representing 30 trees, and you have tested 65 participants representing 21 family trees, a quick calculation shows that you have tested 18 percent of the males and 70 percent of the family trees.

With this situation, as you bring in more participants representing the remaining family trees, the geographical location of the progenitors of the family trees might give you insight into the expectations of a match. For example, if all the trees tested to date have the tree progenitors located in Gloucestershire, Wiltshire, Somerset, and Devon, and you test a tree whose progenitor is in Yorkshire — and surname distribution maps indicate an origin there — it would be highly likely that the new participant will bring a new result to the project (though there is no guarantee — there could have been a migration of that distance).

The exciting and interesting part of DNA testing is that it provides information not in the paper records. This enables you to make discoveries not otherwise possible.

When your project is just starting and you are getting the first few participants, so often the expectation is that they will all match. So two trees test out of 50 family trees, and when there isn't a match, I have seen the conclusion that there must be a NPE (non-paternal event), also called a mis-attributed paternity. This term is referring to some event that broke the link between the Y-chromosome and the surname, such as voluntary name change, illegitimacy where the male child took the mother's surname, infidelity, or adoption.

Realistically, when a small number of the family trees, or a small number of the population of the surname have tested, it is premature to declare NPE events with no evidence. The odds are that down the road, as more of the surnames test, a match will be found.

Let's take a surname with tens of thousands of males today, which is clearly a multiple-origin surname, and the project has 250 participants (less than 1 percent of the population). Is it really reasonable to assume that if you have the surname and don't have a match, then it must be an NPE?

Often this conclusion is based on no matches with the surname, and matches with one or more other surnames.

The problem is that as more of the world population tests, everyone will have multiple matches with other surnames. When surnames were adopted, multiple men had the same or close Y-DNA result. They adopted different surnames. This could lead to matches today. Convergence is another factor that can lead to other surname matches. Convergence is where results mutate over time so they are closer today than before.

Your documented trees tell you about the trees from the start of the documentation until today. But we still have an approximately 400- to 500-year period prior to the documented trees, where we have genetic trees. What transpired during that 500-year period could be significantly different from what transpired during the period of documented trees.

In going back, most documented trees end in the 1800s, 1700s, or 1600s, depending on the surviving records for the tree. If we use the date range 1250-1450 as the adoption of surnames, and take the midpoint as 1350, then a 400-year period takes us to 1750.

Let's take a situation where there are four surname origins. One origin ramified well in the early years of the surname, and many males were created. This event gave this origin a head start in creating a large population today. Two genetic trees had limited males, and one tree had only one surviving male per generation who had males until 1700, when there

were five males whose branches later ramified well, with one branch which went to colonial USA and had 10 sons establishing quite a large future population in the USA.

You start a DNA project for this surname. You test two participants, each representing a different tree. What are the odds of a match? Early in a DNA project the odds are low for multiple-origin surnames. The odds increase as more trees test.

Surname distribution maps, such as from Surname Atlas, are very helpful in providing information about your surname. Extracting and mapping the UK 1851 census with GenMap will enable you to look at the surname a little further back in time. Collecting early recordings and mapping them is helpful. Extracting all online parish registers and mapping them is also helpful.

If your surname is locative and comes from one location, you can have multiple different DNA results and one surname origin. This will occur when multiple men moved away, perhaps in different directions, and became de XYZ, and later XYZ. They could have easily had different DNA results at the time of adoption of surnames. All the maps indicate a single origin, yet you test the trees and the trees don't match.

It is almost impossible to predict what a participant will bring to a project, except if you are testing in a documented tree that has already tested, and even in that situation there can be a surprise or a mistaken connection. The higher the percentage of trees tested, the higher the odds that a new participant representing a new tree will match someone in the project.

## Validation

When you start your DNA project, the ideal situation is to test two distant males for each family tree. These two participants will most likely be an exact or close match, validating two major branches of a tree.

It is often difficult to find the ideal two participants for a tree, when they need to pay for their test kit. Donations may help in this situation. For example, you want to test the male in a branch with the fewest generations back to the progenitor, to minimize the risk of mutations that make your analysis more difficult. Sometimes this man doesn't want to participate, and the one who wants to pay to test is the grandson, adding two more generations with potential of mutations.

Moving forward, you have your two participants, and their results are back from the lab. The best case is that the these two participants are an exact match to each other.

That doesn't always occur.

Let's take a situation where the first two results differ by a genetic distance of one. Here is an example:

**Results Markers 1-18**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant B | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 12 | 13 | 13 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 |
| Participant C | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 12 | 13 | 13 | 29 | 18 | 9 | 9 | 11 | 11 | 25 | 15 |

**Results Markers 19-37**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant B | 19 | 30 | 15 | 15 | 17 | 18 | 11 | 11 | 19 | 23 | 16 | 14 | 16 | 18 | 37 | 38 | 12 | 12 | |
| Participant C | 19 | 30 | 15 | 15 | 17 | 18 | 11 | 11 | 19 | 23 | 16 | 14 | 16 | 18 | 37 | 38 | 12 | 12 | |

We don't know at this point which result is the mutation, whether it is the 17, which is colored yellow, or the 18, the value of the marker directly below the 17. So one of the two markers that differ is selected as the mutation, and highlighted yellow.

There are two ways to figure out which is the mutation.

- You test a third male on this tree to determine the ancestral result.
- As you test other trees, you find several tree-to-tree matches, and the ancestral result will become clear.

In the case under discussion, we tested a third male. All three males descend from a man born circa 1680, each representing a different son. We can determine which are the mutations with this approach, since on the markers that differ, two of the men always match. Therefore we know the ancestral result, which is the third line of the chart. The mutations are now colored grey, since we wouldn't be concerned with them anymore, and any tree-to-tree matches can be analyzed using the ancestral result.

**Results Markers 1-18**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant A | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 12 | 13 | 13 | 29 | 18 | 9 | 9 | 11 | 11 | 25 | 15 |
| Participant B | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 12 | 13 | 13 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 |
| Participant C | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 12 | 13 | 13 | 29 | 18 | 9 | 9 | 11 | 11 | 25 | 15 |

**Results Markers 19-37**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant A | 19 | 30 | 15 | 15 | 16 | 17 | 11 | 11 | 19 | 23 | 16 | 14 | 16 | 18 | 37 | 38 | 12 | 12 | |
| Participant B | 19 | 30 | 15 | 15 | 17 | 18 | 11 | 11 | 19 | 23 | 16 | 14 | 16 | 18 | 37 | 38 | 12 | 12 | |
| Participant C | 19 | 30 | 15 | 15 | 17 | 18 | 11 | 11 | 19 | 23 | 16 | 14 | 16 | 18 | 37 | 38 | 12 | 12 | |

Please note that for the first man, the mutation is occurring in a multi-copy marker, DYS 464. This marker typically has four copies. If it has more or less, a result for each copy is reported by Family Tree DNA. This marker is scored differently. The results are reported from low to high, and this is *not necessarily* their actual order. (If you want to know their exact order, to resolve a situation with two or more participants with mutations on this multi-copy marker, order the DYS 464x test. This test reports the result for this marker in the actual order.)

Let's look at this marker for a moment. The bottom two men have 15 15 17 18 reported for DYS 464. The first man has 15 15 16 17 reported. If we line them up and compare, the first man has a 16, but he also has a 17 matching the other men. This is the reason the 17 isn't highlighted, although it doesn't match what is directly below it, is that it matches "within" the marker copies. Looking at this marker alone, as shown in the chart below, makes it easier to see the situation.

| DYS 464 |
|---|
| 15 15 16 17 |
| 15 15 17 18 |
| 15 15 17 18 |

## Tree-to-Tree Matches and Genetic Groups

When you start having tree-to-tree matches, if the results are exact or close, you can often skip testing the second male for a tree. The closeness of the results enables you to delete the validation step in these situations.

**Y-DNA Results**

| Tree | Result | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 10 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 12 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 17 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 12 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 17 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 24 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 39 | 12 | 12 |
| 41 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 40 | 12 | 12 |
| 41 | 13 | 24 | 14 | 10 | 11 | 14 | 12 | 12 | 12 | 13 | 12 | 29 | 17 | 9 | 9 | 11 | 11 | 25 | 15 | 19 | 29 | 15 | 16 | 16 | 17 | 11 | 11 | 19 | 23 | 15 | 15 | 18 | 16 | 39 | 40 | 12 | 12 |

**Legend:** This chart shows Y-DNA results for a group of family trees that match. The yellow markers are the mutations that have occurred. Grey is a resolved mutation that occurred since the progenitor of the tree, so it is not factored into the analysis. A mutation is nothing negative, simply the scientists' name for a change which happens randomly. These men share a common ancestor between the adoption of surnames and the start of the documented trees, since they share a surname, and no documented connection can be found between the trees. Most trees end, going back, in the 1800s or 1700s, and some end earlier in time. The adoption of surnames was a process that occurred from 1250-1450 in England, so Y-DNA is also spanning the time period from the adoption of the surname to the start of the documented trees, a period of hundreds of years.

In the chart above, two men tested from tree 10, validating that tree. The first man who tested from tree 12 matched, and based on both the match and the fact that all branches were well researched, no validation was needed. For tree 17, two men tested, and one had a mutation. A second man was tested, primarily since the confidence in the research of a branch was low, as well as due to the mutation.

Tree 24 tested one man. And two men tested for tree 41, and these two men both share the same mutation.

Each of these documented trees has a different progenitor, and you now know the ancestral result for each tree.

These five documented trees represent a genetic group. It is recommended that you organize your DNA results at your DNA website at Family Tree DNA into genetic groups using the subgrouping tool. (Log into your GAP, click Project Administration, click Member Subgrouping.) You can give genetic groups titles that will display. There is a description field that you can fill in for each genetic group, but it does NOT display at your public website or your results charts, so it is of limited value.

## Haplogroup R1b

Interpreting results in haplogroup R1b is the most difficult, due to the size of the population (comprising over 45 percent of Europe, and higher in areas in the British Isles), and the fact that since the men are all in the same haplogroup, they are therefore related at some point in time, either prior to the adoption of surnames or in a genetic tree. (If they are in a documented tree, you know when they are related.) Your challenge, when you have a genetic distance of three or four between two men in different documented trees, is to determine if the relationship is in a genealogical time frame — or if they are related prior to the adoption of surnames, and the fact that they share the same surname is a coincidence.

Often validation is useful in this situation, since it will either bring the two trees closer together (indicating a genealogical time frame) or you will find more mutations with the validation partner, increasing the genetic distance.

## Genetic Trees vs Documented Trees

With your research, you create documented trees which end at some point going back in time. With DNA testing, you uncover information not available in the paper records. This may result in identifying mistaken connections where you connected a man to the wrong tree. This can easily happen in an area with a high population of a surname, or upon migration. Validation testing helps to find these situations.

You then group together trees that are exact or a close match into genetic groups.

A genetic group could indicate a surname origin or an event — such as illegitimacy and a male child taking the mother's surname — prior to the start of the documented trees. Collecting early recordings, combined with surname distribution mapping at various time frames, may help in determining which of the two situations has occurred.

## Recruiting

It all starts with recruiting. To make discoveries, you need participants. One approach that works to recruit participants is to stress the discoveries they can make: about their family tree, about their surname, and about their distant origin. They are also making a contribution to the knowledge about their surname.

If your one-name study includes constructing trees, this provides you with a powerful advantage for recruiting. You can avoid recruiting more participants than you need for a robust tree, and focus on recruiting those trees with limited surviving males.

In addition, a very helpful tool for recruiting and providing information about the surname is to build a chart showing the current population by country of each surname in your ONS, along with the earliest documented event for your surname for the country. This tool is an excellent guide for recruiting, especially when all the trees aren't constructed yet.

## Want to Get Started?

When you are ready to add DNA to your one-name study, the DNA Advisor is here to help, including setting up your project with proven marketing material. Simply write: DNA@one-name.org. You will receive a completely setup project that you can modify, along with an easy to follow 20-step Getting Started email and a sample recruiting email and letter. ■