Thoughts On DNA Testing For One-Namers

by David Bicket (4268)

Introduction

This article is intended as a self-help guide for individuals running a surname study who want to get good value from DNA testing in that study, taking advantage of current scientific knowledge about Y-DNA testing. This article is not intended for individuals with a general interest in DNA testing, nor is it a guide for people trying to understand the results of their tests.

This article focuses in particular on the benefits which can be obtained from advances in the last several years related to Y-SNPs. This does not necessarily require Y-SNP testing, because Y-SNPs can be predicted from Y-STR results. But Y-SNP testing will give the most detailed and most usable results. In any case, using this article requires a reasonable understanding of DNA and DNA testing, especially of Y-STR and Y-SNP testing. Annex A (Summary of DNA basics for one-namers) provides a brief overview. All types of DNA testing for genealogy are briefly covered, but the focus is on Y-DNA testing since this is the most relevant for surname studies. By way of explanation, there are significant differences between what is needed to understand autosomal DNA testing (the most widely used in genealogy), and Y-DNA testing. Within Y-DNA testing, there are significant differences between what is needed to understand Y-STR testing and Y-SNP testing. It is not possible for the reader to appreciate the objective of this article without understanding these differences.

This article presents the consolidated experience and opinions of a number of one-namers who have Big Y 700 or equivalent experience, but does not claim to be definitive. The remaining errors, omissions, and biases are purely those of the author. Furthermore, the state of the art in DNA testing continues to advance, and as a result what is appropriate guidance for one-namers will likely change over time. This will happen especially as WGS testing becomes more widespread, which will someday probably be the standard test for everyone, and as more SNPs are identified and mapped into the Y-haplotree.

This article focuses on Y-DNA testing from Family Tree DNA (FTDNA), but does briefly address other types of testing, and other companies. Within the community of FTDNA users, there has been a strong tradition of the use of STR testing, and this article therefore reflects that situation. However, there is also a smaller community of users of YSEQ/YFULL, where more knowledge is required, but where the focus is on WGS testing and Y-SNPs almost exclusively, and Y-STRs are seen as inappropriate to pursue from multiple perspectives.

There is no 'one-size-fits-all' set of recommendations because the situation for any surname study will have its own characteristics which help determine an appropriate course of action for DNA testing. Some of the specific factors that help determine an appropriate course of action are:

How many unique origins your surname has. If the surname of your study has common origins such as occupational names, location names, or 'son of' names, then there are likely to be large numbers of independent origins for people with that surname (and its variants). It will almost certainly not be meaningful to establish the relationships between such independently originating lineages, because they will usually be long before genealogical timeframes. In these cases, many one-namers will consider it sufficient to identify the groupings of different origins, and STR testing is well suited for this.

What are the interests of living individuals with your study surname. Why do people want to join your project and do DNA tests for you? Simple curiosity? Looking for a way to break down their brick walls? Finding where they came from in terms of location and culture? Some may be interested, but not if they have to pay. Some may be willing to pay, but want the cheapest options to address their interests. Some may be willing to go for the most expensive options, and to pay yet more to support tests by others. These factors may help determine the types of DNA testing you promote.

Whether you care about determining the exact relationships between different lineages of a tree which are genetically related within a genealogical timeframe. Another way of expressing this is whether you care about determining the exact relationships between all of the documented lineages of an STR grouping, which will have been within a genealogical timeframe. Perhaps you care primarily about your own STR grouping, but not the many other STR groupings in your surname project.

How many existing STR tests you have in your surname project. If your surname DNA project is quite new, and has few existing testers, then you can follow a course of action designed to get maximum benefits from whatever financial resources you have. If your surname DNA project already has large numbers of test results, most likely at Y-37 or less, then there are similar opportunities to get maximum benefits from whatever additional funds you have to spend. But there may be inertia in the existing approach which could be hard to change.

What is sufficient to satisfy your own curiosity. How much time and money are you willing to spend to satisfy your curiosity? Or to organize others to do the same?

Summary

The guidance in this article can be summarized as follows:

Y-STR uses. Y-STR testing has many potential uses, the most significant of which is to establish groupings of testers who are almost certainly part of the same DNA lineage within genealogical timeframes. Likewise, they can demonstrate that different groupings are almost certainly not related within genealogical timeframes, and that they therefore have independent origins.

Y-SNP uses. Because of significant advances in Y-DNA testing over the past several years, the detailed results provided by SNP testing are now the best available for a surname project. SNP results provide everything which STR results can provide, and much more, because they show where testers are connected to the tree of mankind (the Y haplotree), and thereby specifically

how they are related to other testers, and not just that they are related by being part of the same grouping. The test from FTDNA which provides the full range of SNP results is the Big Y 700 test. (It also provides STR results, but these are incidental, and of limited relevance given the value of the SNP results which are provided.)

Inhibitors to Y-SNP uptake. This being said, there are two main inhibitors to a mass move to SNP testing. These are (1) the unit cost of the Big Y test; and (2) the large number of STR test results which have already been performed in many surname projects. This article will give some recommendations for dealing with both of these issues.

Starting with Y-37 vs Big Y. It is noted that the Y-37 STR test is described as 'a good starting point', including by FTDNA itself. A more accurate position is more nuanced than that. Always starting with Y-37 can be wasteful both in terms of time and cost, although not by excessive amounts. There is no reason to delay going directly to Big Y testing if testers have been chosen according to strict criteria. It is, however, financially inappropriate to test everyone using Big Y, just as it is financially inappropriate to test everyone using Y-37, when relationships between potential testers are already known. For the purposes of getting meaningful results, in most cases fewer Big Y tests are needed than the number of STR tests which are typically performed. Furthermore, Big Y produces qualitatively much better results, meaning that the 'bang for buck' achieved with Big Y testing can be much better than with just Y-STR testing.

Mass participation. There are related issues of mass participation and funding. It can be argued that the cheaper Y-STR tests are better at encouraging a greater level of participation from people with the targeted surnames, potentially resulting in a more active surname project and likewise resulting in greater funding from those participants. The readers of this article can make their own decisions about such issues. However, from the perspective of DNA results, mass participation using Y-STR tests cannot produce the definitive results of targeted SNP testing, regardless of the number of Y-STR tests performed.

Where are we and how did we get here?

The science and best way of using DNA testing in a surname study has changed as testing technology has advanced. (See Annex A - Summary of DNA basics for one-namers.)

Although people started around 2000 experimenting with various forms of DNA testing, Y-STR testing was quickly viewed, in general, as the best type of DNA testing available for surname studies, and this continued to be the situation through about 2016, and particularly since 2019 when FTDNA introduced their Big Y 700 test and their dynamic haplotree. Consequently, surname projects wishing to benefit from DNA testing throughout this earlier period used Y-STR testing for its maximum possible benefits, which generally involved testing as many individuals with the relevant surname and variants as possible. From these results, it was possible to identify groupings of testers with similar STR marker patterns. These groupings showed broad patterns of surname development, e.g., often indicating different geographic origins for the same surname. However, while Y-STR testing can show that these different groups were not related to each other within historical timeframes, it does not show how people within those groupings are related to each other. Until the advent of SNP testing, showing such connections was only possible by researching historical records, to the extent that such records existed and provided credible evidence for linkages.

Since approximately 2016, Y-SNP testing has increasingly enabled the determination of the exact genetic relationships between different testers, whether they are from different STR groupings, or whether they are from within the same STR grouping. Indeed, Y-SNP testing can bypass the need for determining STR-type groupings of testers, because it definitively shows how testers are related if it is before the surname era (and hence they would be in different STR groupings), or how they are related within the surname era (and hence they would be in the same STR grouping).

Some one-namers may wish to continue the STR approach which existed before 2016, when the main purpose of DNA testing for surname studies was to use Y-STR tests to identify groupings of testers for the purpose of identifying broad patterns of surname development. However, that approach has been overtaken by the developing science and use of Y-SNP testing. A more ambitious purpose for DNA testing for surname studies may today be defined as determining the exact genetic connections between lineages which cannot be determined from historical records.

Need for lineages documented by historical records

DNA testing for surname studies, and in particular Y-DNA testing, does not exist in a vacuum. It exists to supplement work on lineages using historical records. While it might theoretically be possible to investigate genealogy purely through DNA, this article assumes that work to create lineages using historical records should always take priority. This article considers that testing done without prior efforts to determine lineages using historical records is likely wasteful financially. It may actually be counterproductive because the lack of meaningful results may discourage and demotivate participants.

For the avoidance of doubt, it might be useful to clarify what is meant in this article by a lineage defined by historical records, especially in the context of a one-name study. This article defines a one-name study lineage as being a named Earliest Known Ancestor of a given surname (or variant), together with all of that person's descendants, particularly male descendants. Ideally this lineage continues down to living men of that surname, although it may terminate earlier for some or all of its branches.

Selecting candidates for Y-DNA testing should always be done with as full knowledge as possible of all potential testers for a given lineage as defined using historical records. The following are considered appropriate:

For Y-STR tests. Typically two or three testers are needed for a given lineage as defined using historical records. They should be chosen to be as distantly related from each other as possible, within that lineage. Using the STR marker values from these testers, it will be possible to assess a modal value, which will be a 'best guess' as to what the STR values were for the Most Recent Common Ancestor of that lineage. Testing more individuals than three, if they are known to be related within the same lineage as defined using historical records, is quite likely wasteful. Nevertheless it is recognised that many new testers lacking a significant documented lineage

derive personal satisfaction from using STR tests to identify or confirm from which grouping of a surname they are paternally descended.

For Y-SNP tests. Only one tester is needed for a given lineage as defined using historical records. It is generally justified to test more than one person in a lineage only if the results from the first tester are significantly unexpected (e.g., connected to individuals with different surnames within a recent historical timeframe, indicating a possible NPE somewhere).

This being said, where lineages cannot be established through historical records, then Y-SNP testing can help determine named connections. For example, there may be multiple birth candidates for the earliest documented ancestor of a specific lineage. Y-SNP testing may indicate which of those candidates it should be.

General recommendations

The following are recommended for all one-namers who want to make the best possible use of DNA testing in their one-name studies:

Determine what you can about the surname's origins and distribution. There are a number of sources of assertions about the origins of certain surnames, and you should at least understand these, even if they are later disproven, or if additional origins are discovered. There are also a number of sources which allow the creation of surname distribution maps, which can be helpful in guiding future work, whether records-based or DNA-based. For example, if a surname has multiple geographical, cultural or etymological origins, then there will almost certainly be multiple independent lineages of that surname.

Remain focused on establishing lineages based on historical records. You should focus in particular on lineages for currently living individuals with your one-name study surname and its variants. If you have not yet done this, try to make contact with such living individuals as a priority. Prior to the internet, a productive approach was to write to people whose names could be found in public directories, e.g., telephone directories. This may still work today. But a more effective way now is probably to search for family trees on genealogy websites whose managers have the desired surname or its variants. Contact these people, and try to get them involved. There are clearly issues with some claimed lineages, especially those on public websites built largely based on hints. All the usual skills and cautions of doing records-based research need to be applied. One issue is that amateur researchers often ignore surname variants, but by being part of a surname study which clearly recognizes name variants, this issue can hopefully be minimized.

Review and be sure you understand the strengths and limitations of the different types of DNA testing. If you started DNA testing prior to 2019, you may need to update your understandings. There is a summary in Annex A of DNA basics for one-namers.

Understand how to interpret the results of the different types of DNA testing. A significant inhibitor to being able to use different types of DNA testing is not understanding how to interpret their results and incorporate their results into one-name studies. See Annex B.

Experiment.[1] Don't discount any DNA testing approach based just on expectations or negative reports. To the extent that you can afford it, try every major type of testing, with every significant company you can. You are probably the best person to experiment on. You will learn in the process. Furthermore, you may find an unexpected result which you might have never found otherwise. (It definitely happens!)

Start small. Once you have decided what you want to achieve in your DNA research for your one-name study, start small so that you understand what you are doing before you launch into a more massive project. One of the best ways of starting is with your immediate family and lineage. If you are changing (or even experimenting with changing) the approach for an existing surname project, you should also start small.

Use the Y-haplotree. Relate all of your Y-DNA results to the Y-haplotree, regardless of the type of Y-DNA testing you perform. (With Y-STR testing, you will get a predicted Y-haplogroup. With Y-SNP testing, you will get a confirmed haplogroup, at a more detailed level.) Determine where your testers, and their branches, fit on the Y-DNA haplotree, and reflect this in how your results are shown.

Make your findings public, without violating personal privacy requirements. It is recommended that you make your findings available as widely as possible, to demonstrate both to members already participating in your surname project, and also to those outside the project, what you are accomplishing. It is recommended you keep your findings updated regularly, so that readers can see your project is not dormant.

Recommendations depending on where you are starting from

Recommendations for what you can do depend to a significant degree on where you are starting.

Established one-name study including both lineages and DNA (mostly Y-37 STR tests). This category reflects the situation for people who have probably been active one-namers for many years, with much or most testing done prior to 2016. This category also reflects the situation for someone inheriting such a study from another one-namer.

- Decide what are your priorities. If your priority is to run a project for which the main objective is the widest level of participation possible, follow the recommendations listed below under 'Recommendations for increased participation'.

- Otherwise, review all of your groupings based on STRs to clarify which have lineages based on historical records. Try to connect in historical records those individuals which have been grouped together. Use the predicted Y-haplogroups, or actual Y-haplogroups if any testers have taken the Big Y test, to structure existing groupings according to the Y-haplotree for those SNP haplogroups. It is particularly recommended to produce a diagram showing where all of your STR groupings fit into the Y-DNA haplotree. (Note that there are several alternative approaches to obtaining predicted Y-haplogroups, including: (1) Family Tree DNA predicted haplogroup with its Y-STR tests; (2) the program NevGen;[2] and (3) the Y-DNA family grouping app.[3]) Then follow the recommendations listed below under 'Recommendations for connection results'.

Established one-name study with lineages but little DNA work. This category reflects the situation for one-namers just embarking on a DNA study to be joined with their records-based one-name study work. They have neither the benefit, nor the baggage, of large quantities of STR tests.

- Set up a surname DNA project if you do not already have one. It is particularly easy to do this with Family Tree DNA with the assistance of the Guild DNA coordinator. See the link in Annex D (Sources of additional information).

- Follow the recommendations above for established one-name studies which already have both extensive lineages and extensive DNA testing results. Given that there are no, or few, existing DNA tests, it is quite possible that the decision will be taken to follow the approach for connection results, rather than for increased participation.

Starters for one-name studies both for lineages and for DNA

- Set up a surname DNA project. It is particularly easy to do this with Family Tree DNA with the assistance of the Guild DNA coordinator. See the link in Annex D (Sources of additional information).

- Focus on finding individuals from different lineages, and on extending their lineages in historical records as much as possible. As you find these individuals, try to get one tester from each lineage to do a Big Y test which will give you the main benefits of the connection approach. Then decide whether you want to continue following only the connection approach, or instead the approach for increased participation.

Recommendations for connection results

Ensure clarity of objectives: determining the connections between separate lineages which are already defined in terms of historical records as far back as possible.

Try to identify at least one Big Y tester per separate lineage.

Get them to test. Depending on how motivated you are for connection results, you may need to pay (or get others from your project to donate to cover the cost). You want to be able to place testers on the Y-haplotree at as detailed a level as possible. Consequently, it is strongly recommended that you plan for them to take a Big Y test. To spread out costs, you might wish to test first with Y-37, and upgrade later when affordable to Big Y. You can rely on the Y-37 test alone to get a predicted Y-haplogroup, but that will not give you the detailed cascade of SNPs which the tester has, including the most detailed 'terminal SNP'.

Establish and promote use of a project General Fund. Some people may wish to donate to help cover the costs of these tests. It is also a convenient way to fund a test with multiple payees.

Recommendations for increased participation

Ensure clarity of objectives: getting more people to participate in your one-name study by getting them to participate in DNA testing.

Try to identify as many living individuals as you can with your study surname or variants.

Invite them to test. In general, the expectation is that they would cover their own costs, because there may not be much benefit from the testing in terms of connecting different lineages. Y-37 STR testing is usually the most appropriate test to advise for the purpose of increased study participation.

Establish and promote use of a project General Fund. Some people may wish to donate so that others who cannot afford full or part payment for their test(s) may participate.

Publicise your analyses and comments and keep these regularly updated. Consider a private website or blog.

Recommendations for further testing

General upgrade considerations. If considering an upgrade from Y-37 STR markers, bear in mind that, though cheaper, an upgrade to 111 markers will yield little of practical value whereas upgrading to a Big Y 700 test will yield much more useful data: 700+ STR markers (including Y-111 for matching purposes), all your named and private SNPs, your place on the Y-haplotree, your nearest matches, and an approximate age of your TMRCA.

Turning private SNPs into named SNPs. If a tester has a large number of private SNPs, get another closely related person to test. Most of their SNPs should be shared, which will have the result of those private SNPs becoming named, and visible on the Y-DNA haplotree. Consider using the 'Rule of Three' to maximize the SNP information you can get from testing a single lineage as established in historical records.[4] This rule has you select someone very closely related to get as many private SNPs named as possible, and then someone distantly related, but still within the same documented lineage, so as to identify likely signature SNPs for the documented lineage.

Investigating possible NPEs. If the SNP results for a tester are unexpected, there is a possibility that an NPE is involved somewhere. The challenge is to determine where in the records-based lineage that the records are incorrect or, if the documented lineages do not go back early enough, to at least try to identify where and when the NPE may have occurred, for example using census data. One possible approach is to test someone else who is in the same documented lineage, but as distantly related as possible. By records-based genealogy and a process of elimination, gradually testing closer and closer to the person with the unexpected result, it may be possible to determine where in the documented lineage the NPE occurred.

Recommendations for documenting your findings

It is recommended that one-namers document their DNA findings concerning the connections between different lineages (which themselves are established through historical records) at least in the following way:

Show haplotree results. Create a diagram which shows the SNPs for the relevant part of the Y-DNA haplotree on the top, or on one side of the diagram. In principle, this should effectively mirror the block tree showing the Big Y results for the project's testers. (You can also include the haplotree branches for people with only predicted Y-haplogroups, based on Y-STR testing, although these may be rather high-level and therefore less helpful in such a diagram.) For usability, the detail of the

chart should be limited to showing the Y- haplogroups/SNPs unique to your testers.

Show lineages aligned to the haplotree. Show the testers' lineages on the bottom, or alternatively on the other side of the diagram, aligned with the relevant terminal SNPs for each tester. It can be simple, e.g., showing only the Most Recent Common Ancestor for that lineage as established in historical records, and then the tester below/aside that. Alternatively, it can show the entire lineage from that MRCA down to the tester, plus others in that lineage. Showing the entire lineage, however, can create an extremely large chart. An example of a chart is given in Annex D, D.7 Examples of documentation of results.

Consider showing haplotree branches back to Adam. It can be instructive to show as a preface to your diagram the predicted or tested SNPs characterising each grouping/branch of the surname on a bespoke but heavily edited haplotree going all the way back to the genetic Adam.

Annex A - Summary of DNA basics for one-namers

A.1. Relevance of DNA test types for one-name studies

How one uses DNA testing depends on the details available from testing, and on how they are presented. These vary over time, and also in cost. There are multiple companies performing the different types of DNA testing relevant for one-name studies.

All types of DNA tests can potentially be useful in a one-name study, in particular to provide hints about where further research in historical records may produce information about connections. Because of this, all types are listed here. However, it is the different types of Y-DNA testing which are most useful in surname studies, so they will be discussed further in subsequent sections.

mtDNA (mitochondrial DNA) testing. These are the least relevant types of DNA tests for one-name studies based on patrilineal male descendancy, although there are some unusual cases where they could be helpful, such as in the case of Richard III of England. But consider trying one for yourself for the experience. Family Tree DNA and YFULL are the only major companies offering stand-alone mtDNA testing.

atDNA (autosomal DNA) testing. These tests are primarily useful for finding close living relatives, including those with the same surname, and possibly NPEs. However, these are limited to identifying relatives connected through relationships at most five generations ago, i.e. they are of little use linking with anyone based on relationships prior to around 1800. If you can afford it, try one yourself with every major testing company, e.g., FTDNA, Ancestry, 23andMe, and MyHeritage. You may be surprised at the results! Ancestry is generally the best to start with because of its huge database for matches, but MyHeritage is often better for European matches, and 23andMe may produce results from people who were interested in its quasi-medical information who do not test with the other companies.

Y-STR (Simple Tandem Repeats) DNA testing. Today these tests are primarily useful for demonstrating if two males are related within the surname era, i.e., that they share a STR grouping, but not how they are related. See also the further explanation below of 'Differences between STR and SNP testing'. Ancestry offered this type of testing at one time, but no longer. It is now primarily offered by FTDNA (where Y-37 is its recommended starter test for Y-DNA testing), and by YSEQ.

Y-SNP (Single Nucleotide Polymorphisms) DNA testing. Usually when we refer to Y-SNP testing, we are referring to Y-SNP discovery testing, i.e., testing many tens of thousands of locations for possible mutations. (See also the next entry for 'pack' or 'panel' Y-SNP testing.) Y-SNP discovery tests are primarily useful for demonstrating exactly how two males, and their respective branches, are genetically related, by showing the genetic point at which they had a common ancestor, and how they have diverged separately since that point. See also the further explanation below of 'Differences between STR and SNP testing'. 23andMe offers some limited information about SNPs, but not discovery tests. The only companies which currently offer meaningful SNP testing are those which offer NGS and WGS testing, as described further below. It is also possible for advanced users to read some Y-SNPs on autosomal test results, but this is not considered a practical way of obtaining Y-SNP information.

Y-SNP Pack or Panel testing. This refers to testing for a limited number of SNPs which are specified in advance, and so contrasts with discovery testing. This is a relatively inexpensive way to confirm if a tester has specific SNPs. A typical use might be to determine exactly where on the existing haplotree to place a tester who is a member of an STR grouping for which there is already a fairly well-defined haplotree, so that its defining SNPs can be specified from a SNP pack/panel test. The limitation is that the tester may have further mutations, which the test does not recognize because of its design, and therefore does not enable the tester to identify his exact place on the haplotree. Both FTDNA and YSEQ offer pack/panel SNP tests.

Y-NGS (Next Generation Sequencing) DNA testing. Next Generation Sequencing refers to DNA analysis by 'sequencing millions of small fragments of DNA in parallel'. 'NGS can be used to sequence entire genomes or constrained to specific area of interest'.[5] This is a testing methodology, which finds both Y-STRs and Y-SNPs.

Big Y 700. This is a branded version of Y-NGS DNA testing offered by FTDNA, which uses 'target enrichment' to try to read certain areas of the Y-DNA and not others. It reads 94% of the SNPs of the current best YSEQ WGS test.[6] It includes Y-STR results for 700 different markers, and Y-SNP results for over 50,000 locations. It is believed to be the most popular NGS Y-DNA test.

WGS (Whole Genome Sequencing) DNA testing. Whole Genome Sequencing refers to DNA analysis which results in the sequencing of (essentially) the whole genome, not just specific areas of interest such as FTDNA's Big Y 700 test. (The only exception to WGS's full coverage can be some hard-to-read areas of DNA, and even these are slowly being conquered by new technologies, in particular nanopore WGS.) The main company currently offering WGS for genealogical purposes is YSEQ. Findings and developments in this type of testing are continuing at a rapid pace. Although it is now possible to obtain WGS tests that are no more expensive than the FTDNA Big Y test, there are disadvantages: there is presently a much smaller data base with which to compare results, the amount of data to handle is vast and cumbersome, and analysis techniques digestible by laymen are still evolving.

A.2. Differences between STR and SNP testing for genealogical purposes

There are significant differences in the usefulness for genealogical purposes of Y-DNA tests for Simple Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs, or 'snips'). There is also considerable historical 'baggage' in their use, or non-use. The original use of STRs was for forensic purposes, e.g., for identifying crime suspects from their DNA. This was the type of Y-DNA testing that was later adapted for genealogical purposes, and there is a large legacy of STR DNA tests which were performed from approximately 2000. SNPs were known about, but the lack of technology for reading them in sufficient detail meant that they were viewed as being only meaningful for understanding 'deep' or pre-historic evolution. However, developments in the last 6 years or so, and especially since the advent of Big Y 700 in 2019, have totally changed the situation for SNPs, as they can now be read in the detail needed to show DNA trees and branching down to current generations.

The following are the characteristics of STR and SNP mutations as currently known which determine their respective uses for genealogical purposes.

Y-STR characteristics and testing approaches and their implications:

- Low test unit costs. STR results are produced by the cheapest FTDNA Y-DNA tests available.

- Strong indicators of being genetically related within the surname era. In theory, the more STR marker counts two men share, the closer they are related; however in practice it is now recognised that this generality can be very misleading.

- Limited number of locations ('markers') tested. When only a limited number of locations are tested, there is significant opportunity for 'false positives' (apparent matches which are not correct) and 'false negatives' (true matches which are not identified). The less markers, the worse. 12 and 25 marker tests were so unreliable that they have now effectively been abandoned. In theory the more markers, the better, but in practice few people get much benefit above 37 markers, including the 111 marker level. Almost no one has found any benefit from the 700 marker level included with FTDNA's Big Y 700 test.

- Frequent back-mutations. STRs frequently back-mutate, meaning that there can be no assurance that current test results accurately reflect the mutations of prior generations.

- Parallel mutations and convergence. Because of the above characteristics of STRs, there can be parallel mutations in unrelated lineages, and likewise mutations in one lineage which result in its converging in its markers with another lineage with which it has no recent common ancestor.

- Conclusions: STRs are good for showing whether two testers are likely to be related within the surname era, and as a consequence they are good for grouping different testers who are all likely to be related within the surname era. However, it is now generally recognized that STRs are highly unreliable for suggesting exactly how testers are related, i.e., how their respective lineages connect. (In technical jargon, they cannot "build an accurate 'phylogenetic' tree, that is the branching tree from a common ancestor, whether a historical person or a prehistoric distant ancestor."[7])

Y-SNP characteristics and testing approaches and their implications:

- High test unit costs. SNP discovery results require the most expensive FTDNA Y-DNA discovery tests available.

- They scientifically demonstrate exact DNA relationships between two testers, including the DNA mutations shared by their most recent common ancestor, and how their DNA lineages have evolved separately since then.

- Over 50,000 discrete locations tested (compared to 700+ STRs).

- Transmitted to all sons, usually without modification, and they rarely back-mutate to their original value.

- Definitive cascades of mutations. Because mutations are transmitted to sons, and because they rarely back-mutate to their original value, each tester has a definitive cascade of mutations which clearly define his place in the Y-DNA haplotree.

- Since SNP mutations are random, and can occur anywhere in the haplotree, the same SNPs can occur in different branches (clades) of the Y-DNA haplotree. However, the fact that SNPs are inherited in definitive cascades means that the same SNP occurring in a different branch of the Y-DNA haplotree does not cause any confusion, since the cumulative cascade of SNPs for each person will be definitive for their own haplotree branch.

- The NGS testing process can result in some locations having few reads, which are therefore considered low quality reads. However, recognizing the cascade of mutations which is intrinsic to SNPs can allow calling poor quality reads.

- While the essentials of SNP testing are quite well established, there continues to be work on some issues, such as:

  - The reading of SNPs in areas of the DNA which are considered unreliable by some researchers and companies.

  - How to deal with SNP insertions and deletions ('Indels').

Cost implications of STR testing compared to SNP testing:

- The cost to get SNP discovery results is higher per test, but much more detailed data is obtained, and in principle only one tester is needed per lineage as established through historical records research.

- This compares to the typical need to get several STR testers per lineage so as to determine the modal STR values of the lineage. (In theory only two STR tests are needed to identify a modal STR values of a surname grouping or branch, and hence distinguish between branches of the surname that have evolved during the surname era, but in practice more are typically obtained.)

- Overall, the total financial cost to get meaningful results from either approach is similar, with SNP results being much more definitive.

A.3. Importance of the Y-haplotree
A haplotree is simply a genetic family tree derived from SNP data. The human haplotree is nature's family tree of humankind. The branches of a haplotree, called haplogroups, are defined by their grouping of single nucleotide mutations, or SNPs. There is a Y-haplotree which shows the male line of descent, and an mt-haplotree, which shows the female line of descent. For surname studies we are primarily concerned with the Y-haplotree.

We do not create the haplotree, but rather we discover it. As our ability to test gets better, and as we get more testers with their Y-DNA testing results for their specific haplogroup, we continue to get an increasingly well-defined view of nature's haplotree, and of our places in it. There can be issues about how much of the Y-DNA we read, and how accurately we can read it. Those are testing methodology issues, and not issues with the haplotree itself. Prior to 2016-2019, the Y-haplotree was so poorly understood, with so little of it discovered, that it was of little use for anything but deep ancestry and ethnology. Since then, however, the situation has dramatically changed. This is because the haplotree is defined by SNPs, and recent advances in SNP testing allow us to understand the haplotree down to present times, and place testers on it.

While it is not a very exciting term, the 'haplotree' is transformative in terms of how it allows genealogists to understand relationships with a precision never before possible.

A.4. Differences between Y-DNA testing companies
Since Y-DNA testing is the most significant for surname studies, this section considers only those companies which offer Y-DNA testing. Companies are listed in order of recommendation, with comments added to each to explain this.

Family Tree DNA (FTDNA). This is the main DNA testing company used by most one-namers, based in Houston, Texas, and now owned by myDNA Life Australia Pty Ltd. Its main advantage is that it has the largest database of both STR and SNP testers available and publishes probably the most comprehensive and up-to-date Y-DNA haplotree. Therefore, its Big Y testers have the greatest likelihood of finding close matches on the Y-DNA haplotree and of determining other testers most closely related to them, whether within the surname era, or before, regardless of the surnames carried by those matches. FTDNA also offers good support facilities for surname projects. Although understanding and working with Y-DNA test results is not simple, FTDNA overall is probably the easiest to use, without demanding that users understand all the intricacies of Y-DNA testing. However, many criticise its website, and some believe it exploits its near monopoly position.

YSEQ. This is the second-most used company for DNA testing for one-namers, founded by someone originally with FTDNA, and based in Berlin, Germany. It offers a large range of tests, but most importantly full WGS, not just limited Y-NGS as does FTDNA with its Big Y 700, and it therefore gives results covering autosomal, full Y, and mitochondrial testing. Its WGS test costs less (as of writing) than FTDNA's more limited Big Y 700. It does not do analysis, which can be done by transferring results to a company called YFULL.

Nebula Genomics. This company founded in 2019 is based in San Francisco (according to Wikipedia) or Miami, Florida (based on the contact location on its website). It offers three WGS tests, although the cheapest 'standard' test is stated by their website to have only .4 coverage. It has impressive credentials based on its own website, but it is not known how well it performs in practice.

Full Genomes Corporation (FGC). This company, based in Rockville, Maryland, offers a range of products including WGS at various coverage levels, and also interpretation of results from other companies. Prices are higher than for YSEQ. They do not ship to the EU or the UK, apparently because of potential GDPR exposure. Most information about them externally appears to be from at latest 2016; the latest short update on their own website appears to be from 2020; and links on their own website are broken. Based on these factors, it is difficult to recommend them, but it may be useful to know about their existence as a company.

Annex B - How to interpret DNA testing results

The results from STR testing and from SNP testing are presented in different ways which reflect the significantly different characteristics of these tests. Many one-namers will be familiar with the way that STR results are presented, how STR 'genetic distances' are calculated, and how suggested matches are shown. It should be pointed out that their implied accuracies typically imply excessive reliability in "matching" tools and underestimate the impact of convergence. More will not be explained here, but explanations can readily be found on-line in many different resources.

It is appropriate, however, to explain briefly how the most important SNP results are shown, because understanding this appears to be one of the hurdles experienced by one-namers who are just starting to work with SNPs and with Big Y 700 testing.

The two key sets of results from SNP testing are (1) the Y-DNA haplotree showing the location of the tester and his definitive cascade of named SNPs; and (2) the tester's private SNPs, also called private variants. Private SNPs are identified by their 7 or 8 digit location in the Y chromosome. They are not yet shared with other testers, and consequently have not yet been given formal SNP names, and therefore do not yet appear on the FTDNA Y-DNA haplotree except shown as a block. (Confusingly there are some cases where SNPs have apparently been named without being shared.) The value of knowing the private SNPs is that you can anticipate these SNPs being named if you are monitoring the private SNPs from multiple testers. This is because it may take Family Tree DNA several weeks to review the results of new tests, and determine that there are new public SNPs to be named. Knowing the number of private SNPs that a tester has also enables their approximate TMRCA to be calculated.

For Family Tree DNA, these two key sets of results are accessed in the following ways:

Y-DNA haplotree. "Big Y" / "Block Tree"

Private SNPs/variants. "Big Y" / "Results" / "Private Variants" / "Show All" for all columns There is an FTDNA presentation by Casimir Roman (in two parts) from RootsTech 2021 giving a more extensive explanation. See Annex D Sources of additional information for the links.

Please note that representations of the Y-DNA haplotree continue to evolve, as additional SNPs are identified, and as new testers add information about the order in which SNP mutations occurred, and branching of the haplotree occurred. In short, SNPs 'continue giving' more and more results as more testers participate, and as testing techniques improve, without existing testers having to do anything other than wait and spot developments relevant to the parts of the haplotree in which they are interested. (FTDNA does not notify users of these developments, so it is up to users to review the haplotree for their results periodically.)

Annex C - Terminology issues

C.1. Generic terminology

Many of the terms used in one-name studies, and in particular in DNA projects, are highly generic, so can be used by different people, in different contexts, to mean different things, which can give rise to misunderstandings. This Annex discusses some of these terms and related issues, in an attempt to encourage clearer usage of terminology in one-name studies.

Many of the most commonly used terms in one-name studies are subject to these issues of generality or ambiguity, including:

- Tree
- Branch
- Haplotree
- Clade
- Group or grouping

There is no usage of any of these terms which is exclusively correct or universally accepted, and to avoid confusion, anyone using any of these terms should make clear what is meant. Specifically, it is recommended to define any of these in terms of its starting individual, whether that be an identified historical figure, or an individual defined purely by DNA. Alternatively, such an individual may be asserted and identified in some other way, e.g., by the location where he or his descendants lived.

Let us start with 'tree' and 'branch'. While the terms are intuitive, they nonetheless can result in significant mis-communication when understood differently. (Interestingly, neither the Family Search nor the Ancestry glossaries define these terms, except in the context of a branch of the Mormon Church.) Any genealogical or DNA tree can be taken backward or forward in time. The term 'branch' is equally ambiguous. Furthermore, branches are 'recursive', i.e., a branch can have many sub-branches, each of which is equally a branch in its own right.

To emphasize this generality of terminology, consider also the term 'haplotree'. In surname studies we most frequently talk about the Y-haplotree as used by organizations like FTDNA and YFULL. But there are two haplotrees, one for Y-DNA, and one for mtDNA. Beyond that, there is the question of where each tree starts. The Y-DNA haplotree is typically understood to start from the 'genetic Adam', i.e., the Most Recent Common Ancestor of all males alive today. However, as geneticists decipher the DNA of extinct branches of mankind, the Y-DNA haplotree of all mankind extends back much further than that genetic Adam. It also has to be recognized that there can be small differences between the representations of haplotrees published by different authorities, e.g. ISOGG, FTDNA, YFull and the YDNA Warehouse. This is because the data defining them is continually increasing, and each representation of the haplotree has slightly different underlying data.

The term 'clade' is similarly generic. The term is actually from the Greek for 'branch', so shares the same level of generic usage as the term 'branch', albeit within the context of discussions about a haplotree.

The prefix 'sub-' is also generic. There is no usage of 'sub-', such as 'sub-tree', 'sub-branch', or 'sub-clade' which is exclusively correct or accepted, and it therefore depends on how the writer means it to be understood. It is incumbent on the writer to make that definition clear.

C.2. Conflicting terminology

The term 'haplotype' can be used in contradictory ways:

The most common understanding for users of STR testing in the context of genetic genealogy, is that haplotypes define the (more or less unique) groupings of testers who have taken STR tests. (FTDNA has a report listing the haplotypes it identifies in a project.)

However, the National Human Genome Research Institute (www.genome.gov) defines haplotype as "a physical grouping of genomic variants (or polymorphisms) that tend to be inherited together. A specific haplotype typically reflects a unique combination of variants that reside near each other on a chromosome." The examples it gives are specifically of SNPs.

The similar term 'Y-haplogroup' applies exclusively to the Y-DNA haplotree. ISOGG defines a haplogroup as "a genetic population group of people who share a common ancestor on the patriline or the matriline. Top-level haplogroups are assigned letters of the alphabet, and deeper refinements consist of additional number and letter combinations. For Y-DNA, a haplogroup may be shown in the long-form nomenclature established by the Y Chromosome Consortium, or it may be expressed in a short-form using a deepest-known single-nucleotide polymorphism (SNP)." As such, there is effectively total overlap in terminology between 'haplogroup', 'clade', and 'branch'. It just depends on what terminology is most appropriate in the context.

C.3. Lineages

The term 'lineage' may be used in different ways, similar to the terms discussed in C.1 above. However, for the purposes of this article, and to avoid ambiguity, this article defines a one-name study lineage as being a named Earliest Known Ancestor of a given surname (or variant), together with all of that person's descendants, particularly male descendants. Ideally this lineage continues down to living men of that surname, although it may terminate earlier for some or all of its branches.

This article most often refers to lineages as established using historical records. See 'Need for lineages documented by historical records' above.

C.4. Surname era and genealogical timeframe

The terms 'surname era' and 'genealogical timeframe' are typically used interchangeably, to designate the period starting when surnames were established, down to the present. However, these terms themselves are somewhat arbitrary, depending on when surnames became established in particular areas, and also on the class of the family. For example, aristocratic and other high-status families typically

had clearly documented lineages long before surnames became established for common people.

In the context of Y-DNA testing, the groupings identified by Y-STR testing are generally understood to identify testers who are related to each other, with a common ancestor, within the surname era, i.e., within a genealogical timeframe. This generalization needs to be tempered with the above comments about the arbitrariness of these terms. Where the groupings identified by Y-STR testing are not always clear, this might indicate higher than expected levels of Y-STR mutations. It might also indicate valid origins for the surname slightly earlier than commonly expected. Y-SNP testing should resolve the issue.

C.5. Naming Y-haplogroups in long-form or short-form nomenclature

As stated above, Y-haplogroups may be described in long-form nomenclature, e.g. R1ba2b3c4, established by the Y Chromosome Consortium, and used in the ISOGG haplotree, or they may be expressed in a short-form using a single-nucleotide polymorphism (SNP), e.g. R-U106. Either is academically correct, but the challenges of using the long-form nomenclature as one gets further down the Y-DNA haplotree make it too cumbersome for anything but discussions of 'ancient' branches of the haplotree.

Annex D Sources of additional information

D.1. Web references

- https://www.genome.gov/genetics-glossary
- https://blog.familytreedna.com/wp-content/uploads/2019/03/big-y-700-white-paper_compressed.pdf
- https://dna-explained.com/2020/05/30/y-dna-step-by-step-big-y-analysis/

D.2. Book references

- '*Tracing Your Ancestors Using DNA*', Pen & Sword Family History, 2019, edited by Graham S. Holton. See particularly chapter 5 'Y-DNA tests' by Alasdair F. Macdonald and John Cleary.
- '*Advanced Genetic Genealogy: Techniques and Case Studies*', Wayne Research, 2019, edited by Debbie Parker Wayne. See particularly chapter 14 'The Promise and Limitations of Genetic Genealogy' by Debbie Kennet.

D.3. Guild help setting up a surname DNA project

- https://one-name.org/want-a-dna-project/ (You must be signed in to access this page.).

D.4. Guild presentations

- https://one-name.org/dna-seminar-2022/. See in particular the following presentations:
  - ONS and Y-STR tests (James Irvine).
  - Strengths and weaknesses of using Y-STRs within Y-DNA projects (John Cleary).
  - ONS and SNP / Big Y tests (James Irvine).
  - Haplotrees, mutation history trees and a peep into the future (John Cleary).

D.5. Other presentations

- Casimir Roman FTDNA presentation on Big Y at RootsTech 2021, in two parts. Please note that some of the details presented are incorrect, but the demonstration of how to use the FTDNA interface is very useful.
  - https://www.familysearch.org/rootstech/session/part-1-how-to-interpret-y-dna-results-a-walk-through-the-big-y
  - https://www.familysearch.org/rootstech/session/part-2-how-to-interpret-y-dna-results-a-walk-through-the-big-y

D.6. Websites

- The Gleason / Gleeson DNA Project (https://gleesondna.blogspot.com/).
- SAPP – Still Another Phylogeny Program (https://www.jdvsite.com/).

D.7. Examples of documentation of results

- https://bicket.one-name.net/showmedia.php?mediaID=84
- www.clanirwin-dna.org

References

1. Some people disagree strongly with this recommendation. You can decide for yourself what you think is worth doing.
2. www.nevgen.org.
3. www.ydnagroupigapp.com.
4. 'Big Y-700 Results and "The Rule of Three"', by Bill Wood, https://www.facebook.com/groups/ftdna.big.y/permalink/2039256342892782/.
5. 'What is next generation sequencing', 28 Aug 2013, National Institute of Health, National Center for Biotechnology Information, National Library of Medicine, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3841808/, accessed 5 June 2022.
6. Source: https://ydna-warehouse.org/benchmarks.
7. p121, 'Tracing Your Ancestors Using DNA', 2019.

David is studying the surname Bicket with variants Becket, Beckett, and Bickett. His registered website can be found at bicket.one-name.net and DNA Project Website at www.familytreedna.com/groups/bicket. David can be contacted at dpbicket@btinternet.com.

Can you spare some time to help with a project to improve the Guild website?

We need members to help in two ways:

- Making it easier to find thing on the website. For instance, we have over 300 videos that need to be put into categories and tagged so that visitors to the site can find them when they search.

- There are many pages that need rewriting and refreshing to make them more appealing and easier to read online. You don't have to be a professional writer but good spelling and grammar and the ability to be concise are desirable attributes!

Even if you can only give a few hours of your time, it could make a real difference.

If you have relevant skills and would be prepared to help, you can find out more by contacting our Production Manager, Melody McKay Burton at:

production-manager@one-name.org