

The Advantages of a Dual DNA/Documentary Approach to Reconstruct the Family Trees of a Surname

Chris Pomery

Abstract

Family history research has traditionally been done primarily through a combination of oral and documentary research. Recently, the development of Y-chromosome DNA tests has created a process so that men, sharing the same surname, can verify whether their previous documentary research has assigned them to the correct tree or direct their future research towards documenting it. This article shows how integrating a collection of surname-defined Y-chromosome DNA results produces advantages for a traditional global surname documentary research project and outlines a methodology for managing a dual approach project that combines DNA tests and documentary research where the primary aim is to reveal the origin(s) of the name. While the discussion that follows focuses on the use of national records relating to England and Wales within the context of a surname originating in one of those countries, the methodology outlined and conclusions reached are broadly applicable to any country where comparable records are available.

Introduction

Family historians who expand their researches beyond the discovery of their immediate family are commonly stimulated by an interest in a particular surname which they seek to place within a wider geographical and temporal context than its connection within their personal family tree. Whole surname documentary studies have, however, typically been confined to surnames that are either of low frequency in the general population or identified with a specific geographical area.

The thinking behind this paper is that a comprehensive documentary reconstruction of most surnames back to the mid-nineteenth century is now an achievable goal in those countries where the key national datasets, such as national censuses and civil registrations, are indexed and online. At the same time, further areas of historical study are opening up as (a) the comparative body of Y-chromosome results grows and (b) the number of fully reconstructed surnames increases.

It will clearly be easier to undertake a documentary reconstruction process for those countries that have published online indexes of their key national records

(such as England, Wales and Scotland) and more difficult in countries which do not provide complete online access (such as both parts of Ireland), where there is no centralised record system (such as France), or where the system of surname transmission is more complex (such as Portugal).

Reviewing the situation with regard to UK-origin surnames, I estimate that there are around 6,000 documented surname reconstruction projects underway worldwide of which around 2,200 are registered with the Guild of One-name Studies ("the Guild").¹ Despite its title, the Guild does not set methodological standards for its members' research, requiring them simply to make a loose commitment to collect data at their own pace and to respond to enquiries from other researchers. I estimate that perhaps only 40% of its members are actively reconstructing all their trees using data from the key national datasets now online.

Taking the Family Tree DNA (FTDNA) figure for registered surnames as a baseline, I estimate that there are around 5,000 unique Y-chromosome surname projects underway worldwide, of which roughly 90% are administered wholly or partly through FTDNA. A quick review of the websites associated with FTDNA-registered projects suggests that few have any active docu-

Address for correspondence: Chris Pomery, DNAresearch@pobox.com

Received: September 1, 2009; accepted: October 8, 2009

¹ The estimate of the number of people conducting whole surname projects without being a Guild member is a guess based purely on anecdotal evidence.

mentary research component or corpus of active researchers. The vast majority of projects are structured simply to aggregate same-surname results to the end of identifying ‘genetic families,’ i.e. those name bearers who share an identical, or near identical, genetic signature along the male line of descent. They generally take a passive approach towards the documentary reconstruction process, allowing DNA testees to upload their documentary research but without undertaking any centralised or consistent updating, coordinating or checking of it. A large proportion of projects focus on reconstructing the trees primarily within an emigrant-receiving country such as the US, or to a lesser extent within an emigrant-donating country such as the UK, but rarely integrating the two.

This observation is not intended as a criticism of those project managers. A neutral way of putting it is that they have out-sourced the function of documentary research to the DNA testees individually and are relying on them to update the DNA project as and when they expand their research. Framing the discussion within this terminology, one function of this paper is to suggest the benefits to DNA project managers firstly of taking an active role in coordinating the updating of this documentary research, and secondly of leading that research process actively themselves at the hub of a dual-approach project.

It is a common feature in both documentary and DNA projects to include additional surnames within their remit. In the case of documentary projects these are generally thought of as lexical ‘variants’ of a core surname, the hypothesis being that collectively they constitute a single ‘name.’ The same thinking can be seen among Y-chromosome DNA projects, though in a more idiosyncratic form and wider purpose. Given the higher

relative prevalence of US-based project administrators, some DNA projects include foreign equivalents, i.e. surnames from emigrant-donating countries that are suspected have been naturalised to their Anglo-Saxon equivalent or nearest homophone.


One cannot compare the headline numbers of documentary and DNA surname projects as one is not comparing like with like. It is in the interest of a DNA project to capture results from an extended pool of surnames and there is no penalty for doing so, while the contrary is true for documentary projects where the workload of data collection and reconstruction is greatly increased when additional variant names are added. One trend is clear: the number of DNA-based surname projects is rising fast while the number of documentary-based projects is barely rising at all.²

In summary, it is plausible that the 6,000 documentary projects might include around 15,000 unique and viable surnames (i.e. they are found in the present day population) while the 5,000 Y-chromosome projects include a somewhat larger number.

At the intersection of the two types of project lie an estimated 250 dual-approach projects where the active documentary reconstruction work and the collation of DNA results are being conducted in parallel. Figures from the Guild of One-Name Studies suggest that around 200 surname DNA projects are led by Guild members. The majority of these projects are hosted at FTDNA and are tagged with the Guild’s logo within the site’s surname listings as shown in Table 1. At present

² Based upon the increase over time in the number of surname projects cited by Family Tree DNA, the largest host of DNA projects, and the stable number of members within the Guild of One-name Studies, the only family history society to focus on surname reconstruction.

Table 1
Example of Surname Project Listings from FTDNA Web Site

Project	Members	Description
...		
Pollock	27	The Pollock DNA Project has been organized for ..
Pomerantz	11	
Pomery 	74	Pomeroy, Pomroy, Pomery, Pummery, Pummeroy,...
Pond/Ponds	9	The Pond/Ponds Family Project seeks to find ...
Pontius	16	The Pontius Family DNA project is designed to...
POOLE	69	For 2009, Wales, Counties Down & Antrim, ...
...		

perhaps 20-25% of these 250 dual-approach projects have collected significant numbers of Y-chromosome results, are actively collecting data from available national datasets, and are actively reconstructing all the trees within their surname(s) on a centralised basis.³

Of the estimated 50-60 active global dual-approach projects only a handful have reported their results in any format. I am not aware that any dual DNA/documentary project has yet been written up and published under conditions of peer review. The vast majority of write ups of Y-chromosome DNA projects exist on websites, generally created by the project administrator. Guild members have access through the Guild website to summaries of project reports designed for documentary project managers for two advanced projects (Pomery, 2008; Creer, 2008).

In addition, interim reports are available for the Cave, Carden, and Pike projects (Cave, 2008; Carden, 2008; Pike, 2008). Many articles have been written on the topic of dual projects, including a three-part series by Susan Meates in the *Journal of One-Name Studies* (Meates, 2006a, 2006b, 2006c), and another on linking a DNA project to an historical study by John Plant published in *Nomina* (Plant, 2005).

The guiding purpose of this paper is to show the benefits of the dual approach being pioneered by these early projects, to demonstrate that it is a feasible route for project managers of both DNA and documentary projects to take, and to describe the future benefits for the family history community as a whole as the results from this kind of dual project start being published.

Given that this paper is predominantly addressed to an audience of DNA project managers, the following discussion is written for the point of view of a DNA project manager working towards setting up a complementary documentary project. Interestingly, there is no comparable publication within the documentary arena where the reverse structure, tailoring the discussion from the point of view of a traditional documentary project manager, can be similarly written up at length.

The following methodology is not suggested as a formal orthodoxy that should be adopted by all dual-approach projects or by all DNA project managers. There is plenty of scope for varying goals among dual-approach projects, for example, in terms of geographical focus and documentary rigour. The 'fishing trip' approach to DNA result collection, adopted by many DNA project managers, is certainly a viable route to take at the outset of a project.

My aim here is simply to describe a methodology that I created ad hoc as new resources became available to me

during the past decade and to offer it as a model for adaptation and refinement by others. I am also interested to find out from other researchers whether their projects have followed the same broad trajectory as my own and whether they have reached the same conclusions that I have about how to organise them.

Documentary Project Methods

The goal of a documentary-led surname project is to reconstruct all its constituent family trees back from the present day to their point of origin. At completion it will reveal the geographical origin(s) of the surname and whether it is headed by a single individual. A fully documented project will also account for illegitimate births and for any other non-standard transmissions of the surname from one generation to the next. Indirectly, such a project can also generate statistical data, including time series data, on a variety of historical and social topics, including the span of 'a generation' and the prevalence of illegitimate births over time. Where the social histories of the individuals have also been reconstructed, this data can support an associated economic and social analysis of change and other historical features.

Documentary projects tend by their nature to focus on the origin of a name. The question they seek to answer is "where did the name-bearers come from?" In the context of an English-origin surname, they tend to focus on researching the origin of the name in England. While documenting the subsequent history of emigrants is an interest the priority is more often to document the origin of the surname in the home country.

A second feature is that its focus will tend towards prioritising the histories of the male name bearers, partly because there are fewer of them and specifically because their life histories have such an impact on the transmission and development of the surname.

The standard methodology behind a documentary reconstruction project is to work back in time, generation by generation, from the present day. Such studies often start with oral and personal sources, focusing on the researcher's immediate family, before expanding to include local and national documentary records.

An initial problem for any documentary project is that there is no single source that provides an accurate list of living name bearers. In the UK context, the best list to work from is the last edition of the national electoral roll prior to 2002, the date when it became possible to opt out one's details from the published version. No pre-2002 list is viewable online, though commercial versions on CD can sometimes be purchased second-hand. Subsequent editions of the public and edited electoral roll, and other up-to-date sources such as the British

³ Figures extrapolated from an online survey conducted within the Guild of One-name Studies in 2007.

Telecom list of land-line phone numbers,⁴ perform markedly less well to create a present-day baseline.

Any baseline list can be checked against the total number or records found in the Office of National Statistics database list in order to estimate the number of present-day name bearers in England and Wales.⁵

The most important set of documentary records in a reconstruction process are the national records of births, marriages and deaths from 1837 to the present day ("civil registration records"). The comprehensiveness of these records is not uniform as only in 1874 did it become a legal requirement for individuals to report events. Common sense suggests that early under-reporting of births was higher than for deaths or marriages.

The public index to these BMD registration records has always supplied additional data that greatly supports the process of linking individual event records together to form the profile of specific persons. For birth records the maiden name of the mother is given; for marriages, the spouse's surname; and for deaths, the date of birth or the age at death. However, this additional data is not currently available across all years back to 1837, its inclusion dating respectively from 1911, 1912 and 1866.

A government-backed online index plans to extend the coverage of this additional data back to 1837, though the project to put this enhanced index online stalled in 2008 and no date to complete and publish it has subsequently been announced. Independent mass indexing projects, both those done by volunteers (such as FreeBMD) and those done by commercial firms, have already created online versions of the original unenhanced government index.

Once the civil registration data back to 1837 has been collected for a surname, a quick linkage process can be undertaken to link sets of records together.

- A death record can be linked to a specific birth record based upon the date of birth or age at death.
- A marriage record can be linked to a birth/death composite record, though with less confidence because the inference is made primarily through the forename match. This process is easier to do using twentieth century records as it became more common for people to have multiple middle names.
- A birth record can be linked to a marriage record in the previous generation by matching the maiden name of the mother to the former name of the spouse given in the marriage index. However, these

are technically two different data items. For example, a mother may on a birth certificate indicate her previous married name rather than her surname by birth just as a woman re-marrying may be recorded in the marriage index under her previous married surname rather than her maiden name.

While recognising that there is no agreed standard for linking record data, this is not the place to discuss problems inherent in assessing and sorting data or how one can objectively choose between alternative event linkage options (which mostly can only be resolved by acquiring further data). While the above linkage process will generate a number of unlinked records, individuals and families, it is robust enough to recreate the majority of the members of the majority of trees in the period from the present day back to 1837. The linkage process can be made much more secure by cross-referencing the individual profiles created using the civil registration records with data from the eight national censuses between 1841 and 1911 all but one of which are now online in their entirety from multiple vendors..

To summarise, by cross-referencing the two sets of primary data, the civil registration and the national census records, it is generally possible to recreate most of the detail in most trees within a surname back to 1841. Present-day data from partial electoral roll and telephone directories can broadly be mapped onto this historical data, though with many gaps. There will certainly be a number of unlinked records, e.g. individuals who die or marry but who appear to have no birth record. And increasingly there are a number of records that appear orphaned due to the inability of the indexes to map the complexities of modern society, e.g. births where the mother has not married into the surname (so no corresponding marriage record can be found), or births where the mother was born a name bearer and where her child has not taken the surname of its father.

While very few documentary surname reconstruction projects using all these data have been written up or published in any form, it is possible now to conclude that:

- it is technically feasible to perform the linkage process described above;
- while the linkage process works more efficiently with low and medium-frequency surnames, it is still broadly effective (though correspondingly more time-consuming to perform) even for some high-frequency surnames;
- the cost of corroborating the basic linkages stimulated by the civil registration indexes starts as low as the subscription fee to a single online provider of UK census data.

⁴ British Telecom phone numbers can be put together using localised surname searches at <http://bt.com>.

⁵ <http://www.taliesin-arleyn.net/names/search.php>.

There are two outputs from a surname linkage process useful within a parallel DNA project.

1. The total number of male individuals available as potential DNA testees is already grouped into a finite number of family trees with a point of origin prior to 1837 (albeit with a rump of unlinked male individuals).
2. A list of potential emigrants has been created, namely those individuals where a birth is recorded among the nineteenth century records but no death record has been found.

The linkage process broadly moves the baseline for the surname back from the present day to the middle of the nineteenth century, a chronological distance of some 170 years or, depending on each tree's history, four to six generations.

The new baseline created for further documentary research reduces the numbers of name-bearers whose ancestry is being tracked back in time as the number of name-bearers for any given surname circa 1840 in the UK will be of the order of one-third the present-day level. While research in the period prior to 1840 will focus on parish records, and benefit from the lack of mobility of families compared to the present day, the number of different ways that a surname is spelled tends to increase. One huge advantage of a parallel DNA project is that has the power to uncover genetic connections among surnames not included in the original documentary research programme, potentially expanding the list of recognisable variants associated with the core surname as it exists in the modern era.

Use Within A DNA Project

A standard surname DNA project, working without significant documentary inputs, is broadly speaking a net ready to collect the result of any male name bearer who wishes to pay for a DNA test. Under this approach, the results collected within the project will be biased in two ways relating to the trees within the surname(s) under study: firstly, towards residents of countries that are more receptive to the benefits of DNA testing (principally the USA), and secondly towards members of already documented trees. The latter observation may seem counter-intuitive; after all, one might expect men who've done no family history research to realise that they have the most to gain by taking a DNA test. Experience from my own project suggests that those men whose family members have already done some research, or who have some inkling that they belong to a particular tree, are the first to pay for a test, thus weighting the results for the surname as a whole towards those trees that have already been documented.

Many DNA projects are set up to confirm the hypothesis that one or more previously researched trees link together as a single tree under a common surname-bearing ancestor. Though this is an excellent starting point, my thesis in this paper is that to discover the origins and structure of a surname the only option is to identify every single tree, gradually reducing their number through repeated iterations of documentary research and DNA testing.

A DNA project possessing the information from a documentary linkage process described above is now in the position to take a targeted approach towards DNA testing its name bearers. Instead of testing anyone ready to pay, or men resident in or associated with defined geographical locations, the project manager can now set out to systematically test one male from each documented tree.

Approaches to potential testees can be organised in different ways. For example, one might prioritise those that appear to have origins outside of the hypothesised area of the surname's origin (on the grounds that the DNA result might link them to a tree within that area), or those that appear to have their origins within the area of origin (to determine whether those which stem from that area have the same DNA signature), or those who appear in the largest trees currently documented (of interest potentially to the largest number of living name bearers).

While one can readily create intermediate goals within a project designed ultimately to reconstruct the trees of an entire surname, the goal of a DNA project for a surname with an origin in, say, England is to build up a matrix of results of individuals living in England and linked to trees with origins in England. This is a vital caveat for any surname study using its DNA results to hypothesise about its origins. Just as the most commonly found haplotype does not necessarily signal the DNA signature of the oldest ancestor, even a calculation based upon the number of historical individuals within trees cannot do so unless these calculations are built exclusively upon data gathered within the country of origin.

The simple reason for this is that two populations, for example in England and the USA, have not faced the same conditions and will have reproduced at different rates. The disparity in the genetic makeup comparing different same-surname populations is the result of differences in their reproductive success (brought about by local conditions) and the operation of the founder effect (whereby a new population starts out as less genetically diverse than the original population). No DNA project I am aware of has formally reported on these effects but anecdotal evidence from several surname projects which I have seen confirms both their presence and influence within surname-led Y-chromosome projects.

An initial aim for a dual documentary/DNA project is to test one man per documented tree. While this might seem a huge task, any active documentary research conducted in parallel will over time consolidate and reduce the total number of trees in the combined project. This process happens most quickly where there is a genuinely iterative research process combining inputs from both the DNA and documentary projects as they arise. When two men are found to have the same DNA signature, this directs the documentary research activity towards finding the common ancestor signalled by the DNA result. In many cases this ancestor will have existed after 1841 or a generation or two prior to it. A regular pattern emerges: over time several small trees coalesce into a single larger tree, which by then will have several DNA tests associated with it through its living descendants, all of whom report a consistently held Y-chromosome result.

A common feature of dual projects is that while some documented trees appear to have no living descendants in the country of origin, the trees do hold men who have emigrated and who have living descendants overseas. In these cases one has no option but to DNA test a male member living outside of the country of origin. Figures from the Pomeroys project suggest this scenario covers around 10-20% of trees in an ongoing reconstruction project.

As projects develop over time and the number of distinct trees is reduced, a new priority will assert itself within the DNA project: to test a minimum of two men per tree. This cross-referencing removes the possibility that the solo DNA haplotype associated with any particular tree does not, in fact, reveal the DNA of the entire tree but merely of that individual's personal line which has been contaminated by external DNA.

With two or more men tested per tree, the single DNA signature identified is not then associated with the individuals who have been tested but with the common male ancestor they share. At this point, one can say that the DNA result is that of a specific historical figure who is located both in time and space. In the case of trees where, because there are no living descendants in the home country, the men being tested live outside it, that specific historical figure can have lived no earlier than the original emigrant.

Any wide scale Y-chromosome surname testing programme will throw up inconsistencies that have to be explained, of which the main one will be the presence of DNA within a tree which is different from the DNA of trees geographically close to it or in the rest of that documented tree.

Particularly in the early stages of a dual project, descendants of different trees that appear to originate in the same geographical area may return markedly different

DNA haplotypes. The question then posed is which scenario is more likely: could they one day be documented within the same tree but one of them hold a different DNA result due to an earlier non-paternity event, or do the different DNA results point towards them being two long-established trees of different and unrelated origins? In this context, non-paternity event includes not only marital infidelity, but also social means of introducing different Y chromosomes into a surname line, e.g. adoption of unrelated children or re-adoption of the surname along the female line. One way to resolve this kind of question is to look at the haplotypes associated with other trees originating in the same geographical area. Do they show a dominant DNA signature, or not? The direct way is to take a second test within each of the apparently conflicting trees. This method will answer the question of whether the initial DNA result should be associated with the individual or the tree. Indeed, repeated use of this method can potentially pinpoint the individual who first carried the different DNA within the tree.

By adopting the above approach, the matrix of DNA results collected will reveal patterns of linkages between trees, and specific historical individuals within those trees, rather than between present-day descendants. The historical content and value of the pool of DNA results is made much richer, and much more useful, by developing it in a coordinated manner.

A key point to note is that as a DNA testing programme develops, the need to use contextual data to help refine the hypotheses associated with the results increases. In summary: the DNA results point to potential linkages, the contextual evidence either supports or conflicts with those hypotheses, but only the documentary evidence demonstrates how the tree is actually put together.

I suspect that dual-approach surname studies will find three broad haplotype frequency patterns – a single haplotype dominant, a pair of roughly equal frequency haplotypes, or no dominant haplotype. As yet we have no data on this issue as so few dual-approach studies have reached the point of completion and are able to report their findings on this point, so at this stage it is moot whether they will reveal a consistent pattern across all surname types, frequencies and geographical regions of origin.

Under the method described above, even though the DNA results can now be interpreted on the level of trees rather than individual living males, one still cannot assert that the modal haplotype – the most frequently found result – indicates the DNA signature of the surname's original founder. The contextual information that is useful at this point is the historical number of name bearers in each tree, not the number alive today. Some trees, it turns out are old and thin, i.e. they have very few living descendants but have old origins, while

others may be fat and young, i.e. they have a large number of descendants within a tree that can be traced no further back than a couple of generations. By counting the total number of historical individuals in the trees associated with each DNA signature, a more balanced picture can emerge when hypothesising about the modal DNA signature.

Pomeroy: Worked Example

Details from an actual project will help to illustrate the points made above.

When the Pomeroy DNA project was launched back in 2000, a documentary project had already been underway for many years and much data, including civil registration index entries, collected. Its findings were, however, patchy. A few trees had been researched back to the 1600s, but most existed as fragments. In many cases, a review of submitted research revealed that for some trees one original piece of research had been passed to us by several different people, and while each had added some additional data none had gone back to check the original material. The unchallenged acceptance and re-use of existing material is a feature of family history research that has increased since the arrival of the web.

The decision to test one man per tree was taken primarily to maximise the opportunity offered within a large scale Y-chromosome testing programme conducted by Professor Bryan Sykes' lab at the University of Oxford, and partly to avoid mass mailing the pool of 825 men whose home addresses we had taken from a complete edition of the electoral roll. Slightly more than 300 trees were identified by using the civil registration data (no on-line census data were available at that time), of which individuals from 51 of the largest trees known at that time were DNA tested. The results revealed two strong DNA signatures, a pattern that remains true after nearly ten years' further research.

After the first results came through in 2001, the targeted approach to identifying DNA testees was extended, and by 2007 broadly speaking we were able to claim that every significant tree of UK origin had at least one DNA result associated with it. By late 2009 the number of documented historic trees has been consolidated down to 49, plus an additional 8 where there are no living descendants in the UK but which may have living descendants outside of it in former colonies. All but one of these 49 trees is documented to an origin prior to 1841. Put simply, we are now quite sure that there are no post-1837 trees, evidenced by UK records, that we do not know about.

The drop in the number of trees from about 300 to 49 appears dramatic, but it was largely achieved by patient documentary research. Back in 2001 we were still

waiting for the first of the national censuses to go online. FreeBMD, the most readily available online source of civil registration records for England and Wales, had barely started (as of September 2009 it hosted more than 173 million unique records). As new datasets became available we incorporated them into our documentary research, to the point where we have now cross-referenced census indexes in all but one case from more than one supplier per census. As it stands, only several hundred civil registration events out of some fifteen thousand post-1837 records remain to be linked to a specific individual and which we do not anticipate will be resolved by the arrival of the eventual enhanced index to the civil registration records.

Critical to developing the dual project approach has been the availability of funds to support simultaneously both mechanisms for resolving doubt or unravelling inconsistencies: further DNA testing and the purchase of documentary data or access to it. Our dual project is backed by a family association willing to subsidise DNA tests where needed. Our documentary research is led from the centre and recorded in a single offline database, though other more distributed or online models of collaboration could work at least as well.

Pomeroy Project Results' Analysis

As things stand, we have now DNA tested the oldest common descendant that we can in almost every tree associated with our main surname and three originally defined variants. Put simply, we've run out of men to DNA test, except in a few specific instances which we consider to be low priorities.

What strikes me is that the pattern of the DNA results is broadly the same as we discovered in 2001. What has changed is my interpretation of it.

The two strong DNA signature haplotypes still stand out, and together are associated with 14 of the 49 trees which contain more than 40% of all historical name bearers born in the UK. A further 9 trees are grouped within three other haplotypes, 20 have an associated DNA result recorded only once within the project and 6 small trees are presently untestable due to the lack of a willing testee.

Some of the 20 singleton results, where the haplotype has been found in only one tree in the project, fall in trees with a documented origin back three or more centuries, while others fall within trees where only one individual has been tested to date and could thus turn out to be individual haplotype anomalies caused by a non-paternity event.

Back in 2001 Professor Sykes confidently told me that the pattern of results we'd produced indicated a multiple origin surname, i.e. that different, unrelated men took

on the name and founded trees at different times and places. I was quite happy with that result at the time as it fitted with my own expectation that most surnames will have multiple origins. Indeed, the reason I first contacted Bryan Sykes was to query his result for his own surname, which is of higher frequency in the UK than Pomeroy by a factor for four or five, and which he had declared to be of single ancestral origin.

Nearly a decade later I look at the combined and expanded DNA and documentary data and wonder if the pattern revealed might suggest a single origin after all. Under this scenario all the differences in the DNA results found in trees other than the two strong modal results could be accounted for as resulting from non-paternity events earlier than their currently documented origin but since the period of the settled formation of surnames, the majority in the period 1400-1650. Even more tantalising is the possibility that the oldest trees heading the pair of strong modal haplotypes could both turn out to contain old non-paternity events as part of a single tree. This might potentially link back to the Norman noble family based in Berry Pomeroy Castle, Devon.

Our surname project has reached the limits of what DNA testing, in its current form, can reveal to help us identify and cluster the trees in our project, but there remains substantial documentary research to be done within the period of parish records, roughly 1550-1840, and prior to it, in order to reveal the full picture of the surname's origins and all its true variants.

Ten Key Learning Points

Following are ten key points for dual-approach projects, the first six related to management and the final four related to the analysis of results.

1. The running of a dual-approach surname project is different from a standard all-comers DNA project. It requires a focus on the country of origin of the surname, both in terms of documentary research and in terms of identifying individuals to take the DNA test.
2. While DNA results of name bearers outside the originating country are of interest within a sub-project covering a defined geographical area, they are of no use in identifying the origin of the surname. There will, however, be occasions where no living descendants of a UK-origin tree can be found except in sub-lines headed by emigrants. In these cases the emigrant's descendants are the only living bearers of the haplotype associated with an historical member of the tree.
3. The project manager of a growing dual-approach project will increasingly wish to subsidise the cost of the DNA tests of new participants as the search for

the right man to take the test becomes more and more specific. Sometimes the hunt will lead around the country, or indeed the world, for the one man whose result will add more meaning to the existing matrix of data.

4. Advancing the documentary reconstruction side also requires funds. Alternative individual profiles suggested by the documentary data will wait to be resolved until funds can be applied to solve them, primarily through purchasing the underlying data behind the civil registration event indicated in the online indexes. A project manager also has to reserve time to liaise with other researchers and to help correct mistakes in their submitted research (which you can be sure that they will have made).
5. There are distinct marketing benefits when running a defined tree-based DNA testing programme; it should be easier to raise money and persuade specific people to take part when you can show them that their personal result fits into a wider plan.
6. The project needs a regularly updated report of its findings both as a record of its status and as a marketing tool.
7. The full benefits of a dual documentary/DNA project are only realised when contextual data is included. DNA results can be sorted into 'genetic families' based upon the raw DNA results alone, but thereafter the analysis of the trees and how they link together is driven by contextual data and documentary evidence. The overall aim of a dual-approach project is to produce a set of combined data that is internally and externally consistent.
8. Early non-paternity events can create a significant DNA haplotype within the pool of surname-wide DNA results. The oldest non-paternity event so far documented in a Pomeroy tree falls in the early 1600s, and the associated haplotype is clearly different from any other found within the project. Well distributed haplotypes within a surname, even where documented in an old tree, can plausibly be hypothesised as very old non-paternity events. In many cases researchers are faced with a choice of origins for individual trees that can be documented back to an early origin but apparently no further: is the individual at the head of the tree its original ancestor, and if so did they take the surname as a conscious choice or were they bequeathed a different genetic heritage through a non-paternity event?
9. The dual project approach gradually removes the need for project managers to use the Time to Most recent Common Ancestor (TMRCA) calculation. As the project develops you will find that you are not asking whether two DNA results are related or

not, which the TMRCA calculation is used to adjudicate, but whether two trees can be linked documentarily or not.

10. The rate of female surname transmission (through unmarried mothers) is rising fast in the UK: the current annual rate found in the civil registration records within my surname group is greater than 30%. Barring any further advances in DNA testing technology, this only increases the need for a complementary documentary and oral research approach within a DNA testing programme.

Conclusions

1. A dual DNA/documentary approach will always produce superior benefits for researchers than a single approach of either type.
2. A surname reconstruction project will combine three methods: oral history, genetic history, and documentary history. The overall aim of a dual documentary/DNA approach is to build a surname project that has a consistent set of DNA results that are explained by the documentary evidence.
3. Documentary reconstruction back to the 1840s for surnames of English or Welsh origin is feasible because the civil registration and census data are now readily available online.
4. Documentary reconstruction is best undertaken for surnames up to a certain frequency. Using the Pomeroy project, with fewer than 2,500 living name bearers in the UK across its four constituent surnames, as a baseline I estimate that tackling surnames up to twice that frequency using the dual approach is feasible.
5. After a surname-wide DNA project is completed, the finishing touches to the dual project, and further advances, will always be supplied by pre-modern documentary evidence.

Next Steps

Within the family history community, given that a single method approach is markedly inferior to a dual project approach, I would like to suggest that:

- DNA project managers can profitably explore collaboration with Guild members, and the body of researchers that they each coordinate, in order
 - to increase the number of dual-approach projects;
 - to outsource the active reconstruction of trees to active researchers.

- For documentary researchers, including Guild members and lineage societies, it is doubtful whether any documentary surname project can be deemed accurate or complete without a corresponding genetic testing project being undertaken to verify the composition of the trees as documented. Even in those cases where documentary researchers do not wish to run a DNA project, it is a simple task to set up a surname DNA project at Family Tree DNA for the purpose of collecting any ad hoc DNA results that you can, and you may later be able to identify a DNA project administrator from among their number.

The discussion in this paper suggests that detailed papers in the following area would benefit both documentary researchers and DNA project managers:

- A list and an analysis of the results of existing dual-approach projects, including an evaluation of how many unique haplotypes have been found per surname and evidence of the different genetic histories of their US and UK populations.
- A detailed write up of a single complete, or near complete, dual-approach surname project, or of a single tree headed by a known ancestor analysed using the dual-approach method.
- A meta review of Y-chromosome surname projects to analyse how often new genetic material enters the male line and how often mutations occur, and to create statistics on the number of unique haplotypes found among descendants holding a surname where the documentary records indicate an unambiguous single origin in a specific county.

Looking forward, the insights outlined in this paper suggest that historians may in future wish to:

- Use databases of public Y-chromosome results to identify links between surnames that have not been hitherto recognised and thus shed light on the process of surname evolution in the medieval period. This analysis would be built upon documentary research to establish the presence of surnames in the same location around the time of surname formation. Some surnames with the same DNA haplotype(s) may be related because they stem from the same localised gene pool, while others will be revealed as linked because they share the same genetic ancestor within the timeframe since surnames became established though a previously unrecognised etymological connection.
- Use databases of public Y-chromosome results to develop a regional DNA analysis, for example by collating and reviewing data on all Cornish-origin and Yorkshire-origin surnames to see what patterns

emerge, or a classification analysis, for example looking at the different types of surnames such as locative-origin names or nickname-origin names. Databases of Irish and Scots clans already exist and reveal how a gene pool can be mapped against a range of family/clan names.⁶

- Develop time series of data within genetically-verified family trees, e.g. illegitimacy rates, infant mortality rates, age at death and the length of a generation.

Acknowledgments

My thanks to Debbie Kennett, John Creer, and Susan Meates for comments on pre-publication drafts of this paper.

Disclosure

Chris Pomery has a commercial contract to promote Family Tree DNA in the UK. The opinions expressed in this article are entirely his own.

Web Resources

Guild of One-Name Studies

<http://www.one-name.org>

⁶ To view data for three projects on Ireland by Dr Brian McEvoy of Trinity College Dublin, and data on the Scottish Clans Project, see Web Resources.

FreeBMD

www.freebmd.org.uk

Irish DNA Data

www.gen.tcd.ie/molpopgen/resources.php

Scottish Clans Project Data

www.scottishdna.net/view.html

References

Carden A (2009) *The Carden DNA Project*.

Pomery C (2008) Using DNA Testing in the Pomeroy Surname Reconstruction Project. (29 June 2008).

Creer J (2008) *Creer DNA Study: How DNA Analysis Has Transformed the Knowledge of a Manx Family's History*.

Plant JS (2005) Modern methods and a controversial surname: Plant. *Nomina*, 28, 115–133.

Meates S (2006a) DNA project has produced discoveries in the Meates One-Name Study not possible with paper records alone. *J One-Name Studies*, 9(1):6-10.

Meates S (2006b) DNA testing of tremendous value in sorting out variants in my one-name study (Part 2). *J One-Name Studies*, 9(2):6-9.

Meates S (2006c) Some tips on establishing a DNA project for your one-name study. Final part of a special series on DNA Projects. *J One-Name Studies*, 9(3):8-10.

Sykes B, Irvén C (2000) Surnames and the Y Chromosome. *Am J Hum Genet*, 66:1417-1419.