# Getting the most from a surname study: semantics, DNA and computer modelling (third edition)

*DR JOHN S PLANT AND PROF RICHARD E PLANT*

## *Abstract*

We here address such questions as: what does a surname mean; is it single origin; and, why do some surnames grow abnormally large? Though most surnames are rare, most people have populous surnames.[1] In this article, we consider in particular the evidence that some frequent surnames could be completely or nearly single origin; this would imply that the whole surname relates to a single family that has grown abnormally large. Some populous surnames have a geographical distribution that might be thought to be consistent with a single origin. As yet, such supposition generally lacks support from adequate DNA evidence.

With the onset of DNA testing, some scientists are becoming more active in surname studies and they might be more reluctant than some traditionalists to infer too much from categories of surname meaning. It has sometimes been argued, for example, that a surname is single-origin if it is locative, or that it is multi-origin if it is occupational. However, King and Jobling (2009)[2] DNA tested forty English surnames and found no statistically significant correlation between the supposed semantic category of a surname and its degree of DNA matching into a single male-line family.

Guided by the empirical evidence, our computer simulations identify various possible reasons for a surname family's unusually prolific growth. In particular, chance is a main factor. Also, overall population growth conditions vary widely between different counties. This can go a long way

[1] In 1881 for example, 90% of the population of England and Wales had the most populous 4% of surnames; and, in 1998, 80% had the 1% most populous. D.K. Tucker (2007) *Surname distribution prints from the GB 1998 Electoral Roll compared with those from other surname distributions*, Nomina, 30, pp. 5-22, esp. 6. Also, D.K. Tucker (2008) *Reaney and Wilson Redux: An Analysis and Comparison with Major English Surname Data Sets,* Nomina, 31, pp. 5-44, esp. 18.

[2] T E King and M A Jobling (2009) *Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames*, Molecular Biology and Evolution, 26(5), pp. 1093-1102.

towards explaining the large population of Plant, which is apparently the second largest single-family contender in the favourable growing conditions of Staffordshire. This surname shows relatively little evidence of a significant living population that stems from origins other than that of its dominant family. The initial semblance that Sykes could be the second largest single-family contender in West Yorkshire is more open to debate, since there might be more substantial other origins. To explain its dominant family, it seems necessary to invoke some exceptional characteristics such as more favourable growth factors for its homeland than those pertaining in the available data for the whole county. Also, for the computer modelling, we consider the effects of additional factors such as polygyny, resilience to plague, Y-chromosome linked fecundity, or an early start to an hereditary surname. Such factors can be beneficial in seeing a family through initial precarious times, sustaining its survival through to a small but real chance of subsequent proliferation in more favourable Industrial Age growth conditions.

## Contents

## Preface to the second edition

This is a development of the first edition,[3] which was first published in May 2012. We have slightly updated the DNA results, their analysis, the treatment of Y-chromosome linked fecundity, and we have added some further findings from medieval records. In particular, we have extended the simulations, giving more detailed attention to the effects of both national and regional population growth factors, as well as developing other possible theoretical explanations of the evidence for the surnames Plant and Sykes for example, which both appear to be dominated in England by an abnormally large single family. We have added a Discussion section and amended the Conclusions as well as adding three further Appendices. The first edition of this article was 40 pages long and we have revised and extended it to 63 pages. We are very grateful to Debbie Kennett for some comments on a near-final draft though the responsibility for the final version is ours.

## Preface to the third edition

This further edition contains relatively few updates and some clarifications to the text of the second edition. In particular, some further discussion has been added to the section concerning whether the main Plant family arrived in its main homeland from elsewhere.

## Introduction

In the past decade, DNA testing has prompted revisions to our understanding of surnames. In some cases, the resulting reappraisals have extended to a surname's supposed meanings.[4]

Meaning has long featured strongly in surname studies, with linguists freely accepting such a device as metonymy when postulating a surname's semantics, often neglecting other considerations arising from a more detailed study of the particular surname. Their discipline's practical limitations, in

[3] http://cogprints/8267

[4] John S Plant (2009) *Surname Studies with Genetics*, DNA Section, Guild of One Name Studies (21 pages). http://cogprints.org/6595

dealing with a very large number of surnames, typically involves the need to look for general linguistic trends amongst families of surnames, which may or may not in truth be semantically related. Sometimes, a surname's supposed meaning has been used to suggest a selected hypothesis for its origins, as for Plant which has been claimed to be occupational and multi-origin in order to explain it large population. A fuller investigation, however, reveals that there are at least four different semantic hypotheses for this surname's origins, with different implications for its development and growth. Despite an assertion in some Surname Dictionaries that Plant simply means a `gardener', it may have come from any of at least three different languages and be locative, occupational, or a nickname.

Some people have taken summary surname meanings to indicate the likely number of geographical origins to a name: for example, an `occupational name' is sometimes said to imply many origins unless, of course, it derives from a particular family's rare adoption of its seminal word(s). Some surnames could perhaps have been coined from a dialect word for a local trade such that, though occupational, they could have arisen with just one, or relatively few origins. Though locative surnames are sometimes taken to have a single origin there are for example, for the much discussed example of the Sykes surname, several minor places called Syke, meaning a ditch or stream. In 1379, men named `del Syke' or `del Sikes' were taxed in six different townships in the West Riding of Yorkshire and once at Butterworth in Lancashire.[5] Most people now consider that it has been overstated that Sykes must be a single-origin surname. Mark Jobling and Turi King[6] pointed to the low-resolution of the initial Sykes DNA study though some have perhaps gone on rather to overstate the significance of that single point. A fuller account of the problems in the initial Sykes study has been outlined by Debbie Kennett.[7] We outline the underlying statistical considerations in this article.

As we will show, the surname Berry is clustered largely around Bury in Lancashire and has a very large, apparently single-source population in 1881 that seems too large for a ready explanation as a single family. It might simply be that Berry is in fact more than one family. In their Surname Dictionary, Reaney and Wilson[8] state that, in Middle English, the words *beri*, *biri*, *buri* were used for a manor-house and they conclude that the surname must often have meant a `servant at a manor house'. Without sound DNA evidence, we cannot be sure that its large surname population does not arise from many families that originated with some such meaning, especially around the particular

---

[5] G Redmonds, T King, and D Hey, *Surnames, DNA, and Family History*, (Oxford University Press, 2011), p.19.

[6] King, T.E. and Jobling, M.A. (2009) *What's in a name? - Y chromosomes, surnames, and the genetic genealogy revolution*, Trends Genet. 25, 351-360.

[7] Debbie Kennett, *The Surname Handbook: A Guide to Family Name Research in the 21st Century* (History Press, October 2012), pp. 154-156.

[8] P H Reaney and R M Wilson, A Dictionary of British Surnames, Second Edition (London, 1976).

place named Bury in Lancashire where they thrived, more so than elsewhere where there are for example other such place-names. It might also be noted that, like Plant, Berry is an English word and as such may have served as a convergence for several similar sounding surnames with originally very different meanings.

Despite some uncertainties, we can look to modern mapping techniques for some clues as to whether a surname might be single- or multi-origin. As supporting evidence, DNA investigations can help to ascertain the male-line relatedness of living individuals and how far a single family, perhaps as one family amongst several for a surname, has spread. With the advance in the popularity of the DNA studies of surnames, there is topical interest in whether chosen English surnames are single-, plural-, or multi-origin, where these terms indicate respectively one, a few, or many distinct origins. A received wisdom is that populous surnames in England are multi-origin[9] and a limited DNA study by King and Jobling (K&J) goes some way to support that.[10] However, it was following a pioneering DNA study of the Sykes surname, over a decade ago, that there formed a widespread contrary view, which survives as a contested contention that at least *some* relatively common surnames might be single- rather than multi-origin.

Surnames that are both populous and single-origin are not the norm amongst surnames and so we consider them as atypical arising from exceptional random chance, possibly along with other favourable circumstances. Where possible, in due course, we extend our approach to the relative mathematical probabilities of unlikely events, so that we can assess the relative merits of various explanations for the apparent occurrence of some populous surnames that appear to be single-origin.

## Geography, DNA, and some surname distribution evidence

The geographical distribution of a surname provides a relatively ready criterion for deciding whether a surname might be single-origin. Ideally, the surname's geographical distribution should be traced back to its origins, perhaps as far back as 800 years ago. However, the data for that early are generally incomplete and unreliable. The available studies have been used to generalise that a surname's geographical distribution often remains largely unchanged until recent times. It is generally held that more widespread mobility has become prevalent only in the last century or so, though we shall not neglect the case for exceptions to this.

[9] David Hey, *Family Names and Family History* (Cambridge University Press, 2000).

[10] T E King and M A Jobling (2009) *Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames*, Molecular Biology and Evolution, 26(5), pp. 1093-1102.

We begin with some surname distribution maps obtained using Steve Archer's 1881 *Surname Atlas* CD. For less populous surnames, the chance migrations of a few individuals are likely to affect a surname's overall geographical distribution significantly. A note of caution is needed even for populous surnames. If their initial population was small, in medieval times, the chance migration of a few could conceivably have affected the name's overall subsequent distribution, as also could substantial migrations from rural areas to cities such as to London and Birmingham. There is evidence, for example, that such migrations occurred in the case of the surname Plant. As a simple template however, we follow others in assuming, at least initially, that peasants in medieval times were generally immobile and, once a surname has become populous, relatively small pockets aside, the migration of a few might not greatly affect the general picture of a surname's distribution by 1881.

## Some considerations pertaining to early surname migration

The general evidence for medieval England indicates that most families remained in their small local area. Widespread ramification in medieval England, conceivably beginning as early as the thirteenth or fourteenth century at the start of surname times, seems to have been restricted to only a few members of the general population. At that time, not quite everyone however was a villein, tied by law to their lord's local land.[11] For example, 10% rising to 20% of the population lived in a town. These percentages and the following numbers are taken from a book by Christopher Dyer.[12] Below the `tenants-in-chief' - i.e. the fifty to eighty Dukes, Earls and Barons - there were around 1,100 knights and 10,000 esquires and gentlemen who held their manors from the few principal land holders; the latter were answerable directly to the King. Mobility was possible, if not normal, for these classes and this extended further down the social scale. The 11,100 sub-tenants of the tenants-in-chief were in the "fighting class" of this society, which had a relatively young age profile, and they had legal status and family connections that gave them influence among their peers and power over their own sub-tenants and bondmen. Also, around 2% of adult males in England were clergymen. The church aristocrats were also substantial land holders. Amongst the so-called `peasants', under the secular and church aristocracy, there were some relatively high status franklins, yeomen, merchants and husbandmen. These were sufficiently free that some might have been moderately mobile.

Such freedom can be ascribed, for example, to some of the recorded early Plants but such evidence for possible mobility is not convincingly frequent. Finally, there were outlaws, as indeed was the case for one early Plant: individuals who were wanted for a crime in one county might flee to another to escape prosecution. Although, as mentioned above, there is evidence, for example, for such

[11] Ian Mortimer, *The Time Traveller's Guide to Medieval England* (London 2009), pp.38-47.

[12] Christopher Dyer, *Making a living in the Middle Ages: the people of Britain 850-1520* (New Haven and London, 2002).

migrations in the case of the surname Plant, the geographical distributions of surname bearers, even by 1881, generally appear to show concentrations that can be taken to be near where the surname originated. Such concentrations are, of course, far less readily discernible for surnames that appear to have had very many origins.

## Distribution and other evidence for Smith

Smith is the most populous surname both in England and in Scotland. It has a population of 422,733, in the 1881 UK census, in which the evidence shows it to have a widespread geographical distribution. Both this distribution (Figure 1) and the DNA testing by King and Jobling (K&J) are compatible with Smith being a multi-origin surname. This finding is in keeping with the received wisdom that this name was adopted by many smiths who were distinctive tradesmen in their own separate localities, of which there were many.

The more uniform shading in the left map in Figure 1 indicates that the name Smith was almost evenly distributed throughout the population of England and Wales whereas the shading in the right-hand-side map incorporates the fact that the Smiths, largely reflecting the total population distribution of all names, were not distributed evenly. It can be argued that the left-hand map is more appropriate for a multi-origin name, to show how it might have arisen almost anywhere throughout a widespread population, whereas the right-hand map might be more appropriate for a single-origin surname contender to indicate the extent to which there had been migration from a single source. Though evidently not indicating a single source, the Smith distribution is affected by a high total population around London and Birmingham, as well as around Lancashire at the height of the Industrial Age in 1881.

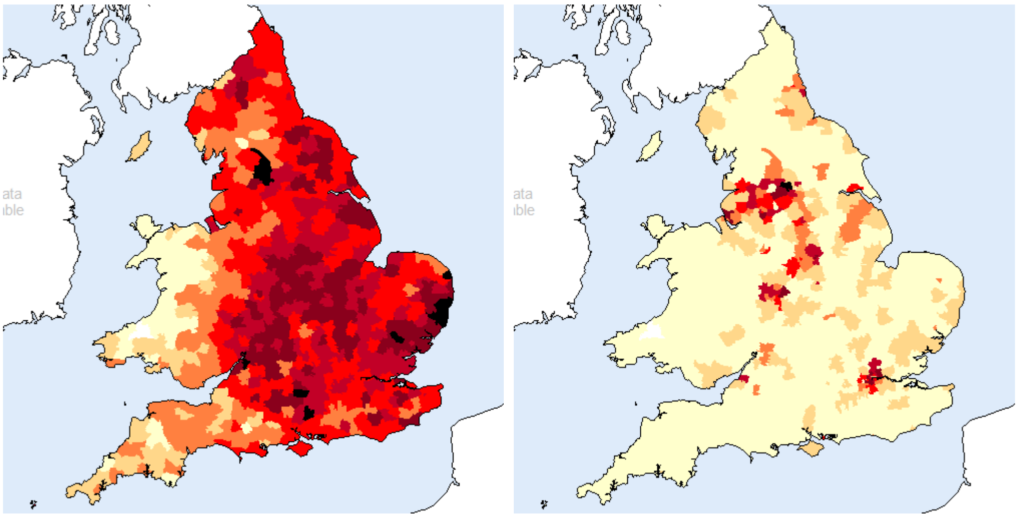


**Figure 1: The distribution of the surname Smith in England and Wales in 1881. On the left, the shading represents the number of Smiths per 100,000 people, for each Poor Law Union, whereas the actual number of Smiths in each Union is represented on the right.**

Turning away for a moment from distributions throughout a geographical map, the concept of “DNA space” might confuse some people and so warrants some words of explanation: this concept relates to plotting out the genetic differences between DNA-tested men and it has nothing directly to do with these men’s geographical distribution. It has more to do with how closely related people are: a small (or large) genetic distance between DNA signatures, within a hypothetical “DNA space”, suggests a close (or distant) familial relationship. Close cousins down the male line tend generally to be genetically closer, in genetic space, than very distant cousins. Clearly, if individuals with the same surname are geographical neighbours they *might be* closely related and hence also close in genetic space; but, clustering in “genetic (or DNA) space” displays familial relationships more directly than for geographical maps, even when individuals have migrated far away.

Turning to the observed clustering in genetic space of the Y-STR signatures for Smith, in the K&J study, the largest cluster of near-matching Y-STR signatures was not a compact one. Put another way, there is only one loosely-connected pattern of Y-STR signatures for the participating Smiths, when differences in these Y-signatures are plotted out. This indicates that those within this small cluster are not especially closely related in modern times. With its diffuse nature, this DNA cluster could be explained largely by an ancient grouping of Y-STR signatures into the diffuse cluster of the R1b1 haplogroup, which is the most common grouping of individuals into an ancient familial relationship that arises in the general UK population; in other words, this Y-STR clustering can be regarded as dating back to pre-surname times since when many mutations have occurred in the ancient family leading to only a loosely-matching cluster around a modal R1b1 signature in the DNA results. Moreover, the whole of this diffuse Y-STR cluster amounted to only 15% of the total number of the Smith men who were DNA tested. K&J accordingly concluded that there was no overwhelming evidence, for the occurrence of a particular single family for Smith, where a single family here means one that is male-line related in the past 800 years, since the formation of this very common surname. That does not preclude that some Smith families might have grown to be larger than others potentially leading to many tighter tiny clusters in DNA space, if many more Smiths were to be tested.

### The less clear-cut situation for Metcalf

As a different example, Redmonds et al have suggested that the surname Metcalf is “single-origin”.[13] It has been pointed out by Debbie Kennett[14] that the DNA results for this surname show six distinct genetic DNA clusters, which might be associated with six different ancestors within surname times. It is relevant, however, to add that this DNA study does not include a discussion of how those tested were selected, leaving some doubt about the extent to which they can be regarded as representing a random sample of the surname. This can have an important bearing on the significance of this Metcalf study’s six mismatching DNA clusters: the clusters might apply partly to disproportionate numbers that have been non-randomly selected for recently-related members of particular families. The six different genetic clusters might represent distinct origins to the surname, rather than a single one, though at least some of the separation into separate cluster might instead have arisen from discontinuities in the male-line descent of the surname by an interposed female passing on the surname. Nonetheless, the available DNA results do *not*, as they stand, support a single-origin contention for Metcalf. This surname has a UK population of 6,065 in the 1881 UK Census and it is widely spread throughout the north of England (Figure 2(a)). Within the extent of the available evidence, neither the 1881 distribution of this surname nor the available DNA evidence supports a single-origin contention for the surname Metcalf.

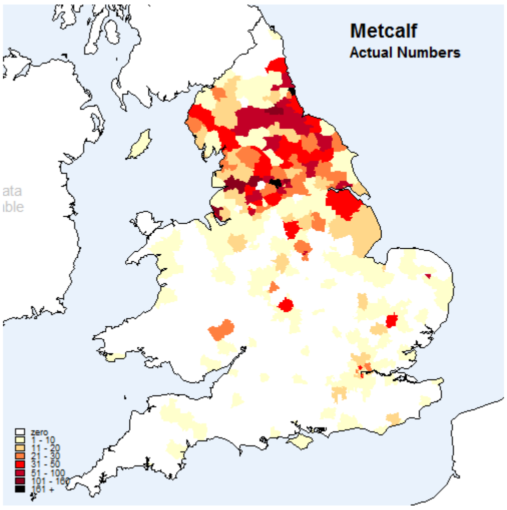
According to the Surname Dictionary of Reaney and Wilson,[15] Metcalf is thought to have derived as a nickname meaning someone who resembles a calf to be fattened for meat. The variant Medcalf is associated with a calf to be so fattened by lush grass in a mead or meadow. Some people consider that such a nickname is sufficiently distinctive to have not arisen very often, implying only a few, if not just one, origin to the surname. As has been outlined however, this is not supported by other available evidence, so far at least.

---

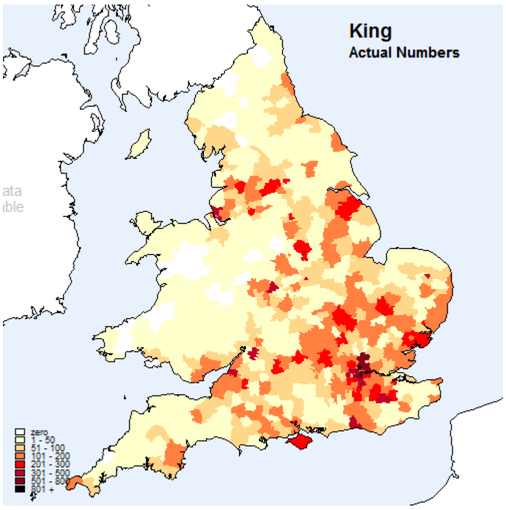
[13] George Redmonds, Turi King and David Hey, *Surnames, DNA & Family History*, (Oxford, 2011), pp. 73-4.

[14] Debbie Kennett, (Jan-Mar 2012), Book Review of *Surnames, DNA & Family History*, Journal of One-Name Studies, Vol. 11, Issue 1, p. 29.

[15] P H Reaney and R M Wilson, *A Dictionary of British Surnames*, 2nd Edition (1976).




(a)




(b)

**Figure 2: The widespread distributions of the two populous surnames, (a) Metcalf and (b) King, in 1881. The shading here represents the actual number bearing the surname in each Poor Law Union, as is also the case for Figures 3, 4, 5, 7, 8 and 9.**

## Available DNA evidence in connection with the distributions of some other surnames

King and Jobling (K&J) considered forty English surnames and these ranged from the populous to the rare. Smith, which has already been mentioned, was the most populous. We here display distribution maps for the next most populous five in their study, each having an 1881 UK population in excess of 4,000.

For the surname King (Figure 2(b)), the UK population in 1881 was 65,233 and this was widely distributed. In the K&J study, its largest DNA cluster was only around 8% of the total number of tested men called King, and this supports the hypothesis that this surname is multi-origin. Reaney and Wilson suggest in their Surname Dictionary that this name originated for those who appeared like, or played the role of, a king; that might have been in a tournament, play or pageant. It is not unreasonable to imagine that such a name could have had a widespread appeal, and such supposition is reinforced by the geographical distribution (Figure 2(b)) and DNA evidence.

Figure 3 shows the 1881 distributions of another four populous English surnames from the K&J study: Bray (UK population in 1881 of 10,040); Stead (6,130); Clare (4,340); and Wadsworth (4,175). The sizes of the largest DNA clusters for both Stead and Wadsworth are around 30% and 35% respectively of the total of those tested; both their DNA matches and their 1881 geographical distributions suggest that these two surnames each include a significant family amongst others. On the other hand, Bray and, to a lesser extent, Clare are more evenly spread and have less DNA matching: their largest DNA clusters are 0% and 25% respectively. This favours rather more a multi- or plural-origin hypothesis for these two surnames on the left side of Figure 3, particularly for Bray. On the right side, it can be noted that Stead and Wadsworth have particular geographical concentrations in the West Riding of Yorkshire where there might have been relatively large families.

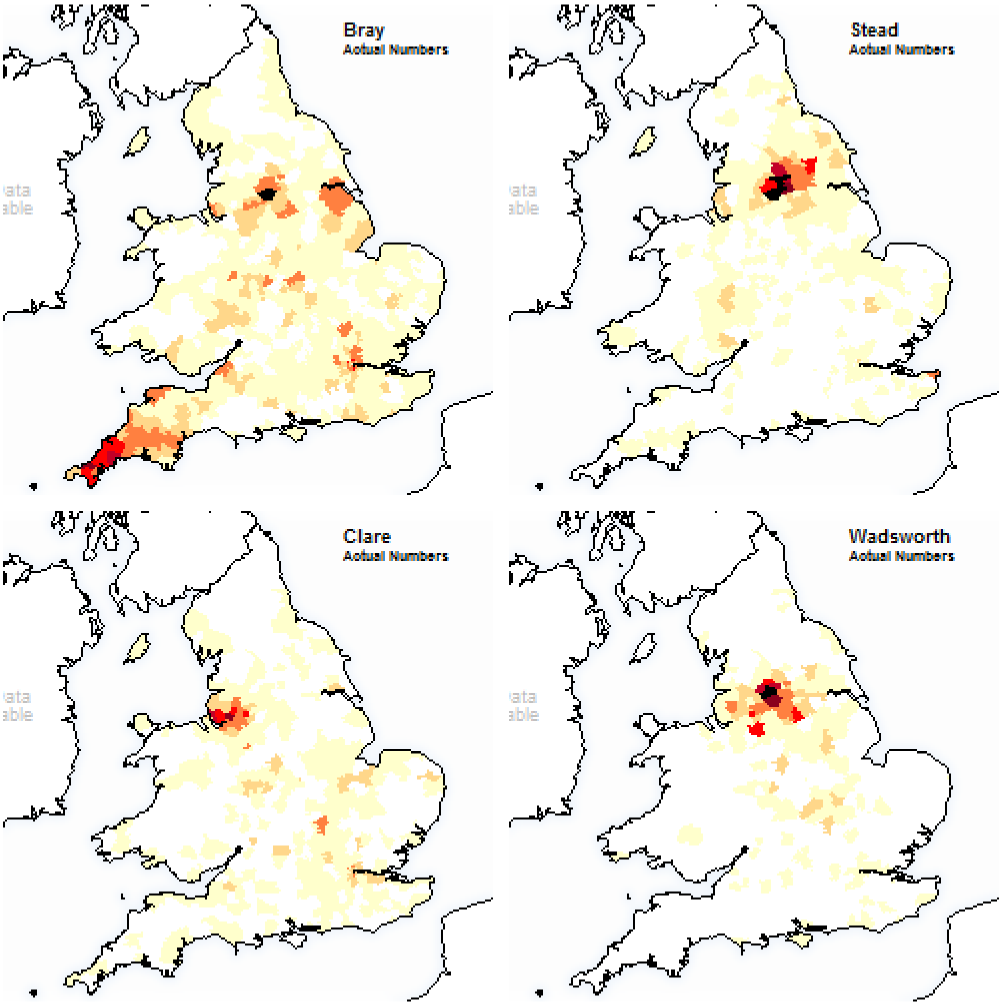



Figure 3: Distributions in 1881 of some populous surnames from the K&J study.

## Populous single-origin contenders in West Yorkshire and Lancashire

Distribution maps for the four most populous single-origin surname contenders in West Yorkshire, in 1881, are shown in Figure 4. These are taken from Steve Archer's Surname Atlas CD. Though Shaw is more populous than these,[16] it is omitted from Figure 4 for being too widely spread. This leaves, for single-origin contenders in the West Riding of Yorkshire: Greenwood; Sykes; Haigh; and, Hirst. These names have West Riding/UK populations in 1881 of respectively: 10,612/22,256; 9,203/14,383; 8,024/10,324; and, 7,646/9,785. The supposition that they contend as single-origin surnames is here based solely on their 1881 distributions; that supposition has not been confirmed by DNA evidence, apart from some supporting evidence for Sykes.

[16] Its West Riding/UK population is 11,979/55,045.

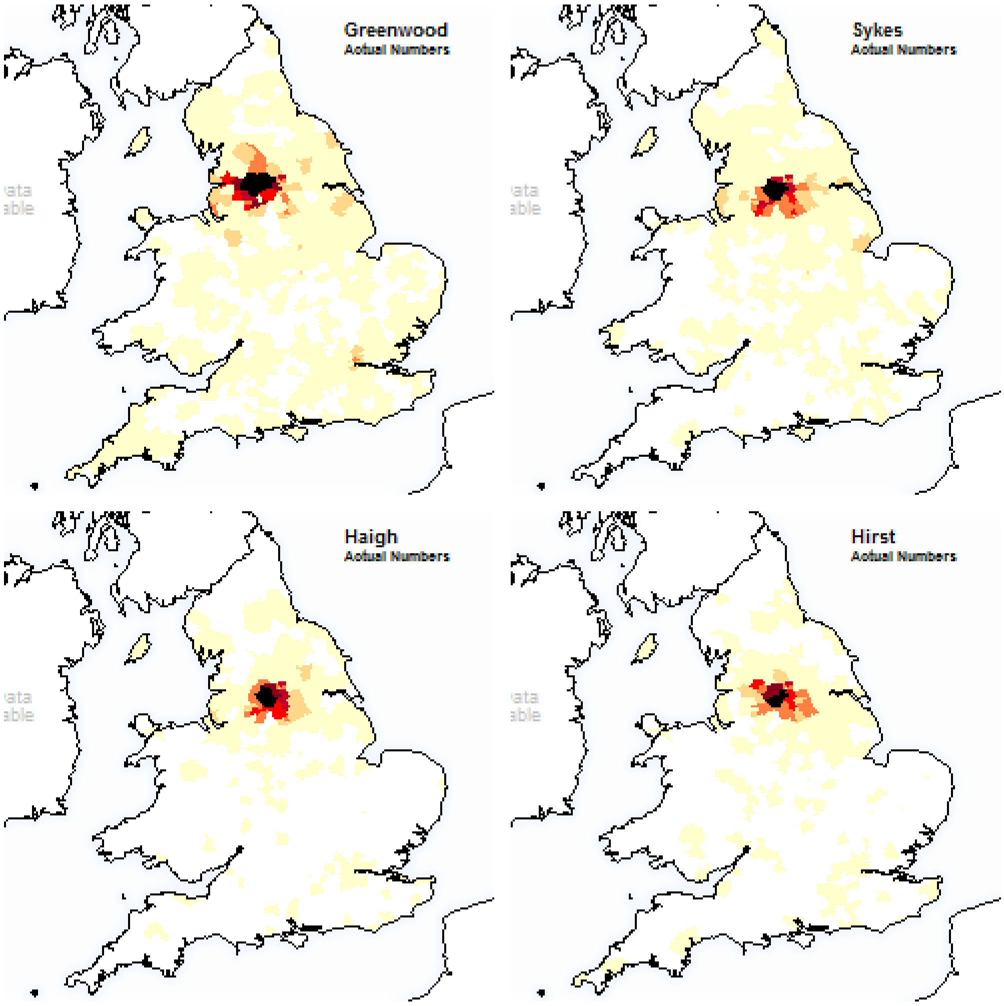



**Figure 4: The four most populous single-origin contenders in the West Riding of Yorkshire.**

The 1881 geographical distribution of Sykes is show in the upper-right map of Figure 4. Though this map was apparently not available to George Redmonds, he took the early DNA study for Sykes[17] as confirmation of his suppositions from early historical records that some populous surnames in West Yorkshire are single-origin.[18] The Sykes DNA study was low resolution and was based on the measurement of only four Y-STR markers for each tested individual. However, the single-origin contention in this study was supported by DNA mismatches to a small control sample of results from non-Sykes neighbours of the tested Sykes men and also mismatches to a wider sample of non-Sykes men. These Sykes surname results displayed a single genetic cluster made up of DNA matches for 43% of the tested men, though statistically this percentage has a 7% uncertainty when extrapolated

---

[17] B Sykes and C Irven (2000) *Surnames and the Y Chromosome*, Am. J. Hum. Gen. 66(4), pp. 1417-1419.

[18] George Redmonds, *Names and History* (2004) pp. 26-31.

to the whole population of the surname.[19] This somewhat surprising DNA result marked a beginning to renewed interest, a little over a decade ago, in whether a single-origin feature might be more common than originally thought amongst populous English surnames. A more recent higher-resolution DNA study for the name Sykes found several large genetic clusters, including the one found in the early English study; but, this American study was both non-random and relied on the descendants of presumably relatively-few emigrants to America who may have had female, instead of male, linkages to an original Sykes family in England – it cannot be ruled out that it may have been by so-called `genetic drift' that female-linked descendants in England went on to father substantial non-matching families growing profusely in more opportune conditions in America.[20] As well as these shortcomings arising from this more-recent largely American DNA study, the early Sykes Y-DNA study for England also has been criticized, such as for sampling the name in only four counties around West Yorkshire. Though the conclusion remains controversial, to say the least, as to whether Sykes is truly a "single origin" name, there does seem some reason to believe from the available evidence that there is a large Sykes family around its main English homeland in West Yorkshire.

For present purposes, later in this article, we will restrict our simulations to family growth conditions in England and then to particular regions of England. We are extending our considerations further, such as to include emigration overseas, in a further paper that is in preparation.

Populous single-origin contenders are found also in the adjoining county of Lancashire (Figure 5). Here, five surnames that are more frequent have been omitted for being widely spread and hence seemingly multi-origin.[21] Omitting these, the largest single-ancestor contenders in Lancashire are: Howarth (Lancashire/UK population is 11,424/14,416); Holt (10,556/20,077); Ashworth (9,554/11,947); and, Holden (9,091/16,421). The largest single-origin contenders in Lancashire are in adjoining regions to those in West Yorkshire.

[19] John S Plant (2009) *Surname Studies with Genetics*, DNA Section, Guild of One Name Studies (21 pages), esp. pp. 3-4. http://cogprints.org/6595

[20] See the Chapter *DNA and Surnames* in *The Surname Handbook: A Guide to Family Name Research in the 21st Century* by Debbie Kennett (History Press, October 2012).

[21] These are Taylor (38,385/191,486); Jackson (18,279/83,702); Harrison (13,389/66,470); Wilkinson (10,308/45,702); Yates (9,188/20,587).

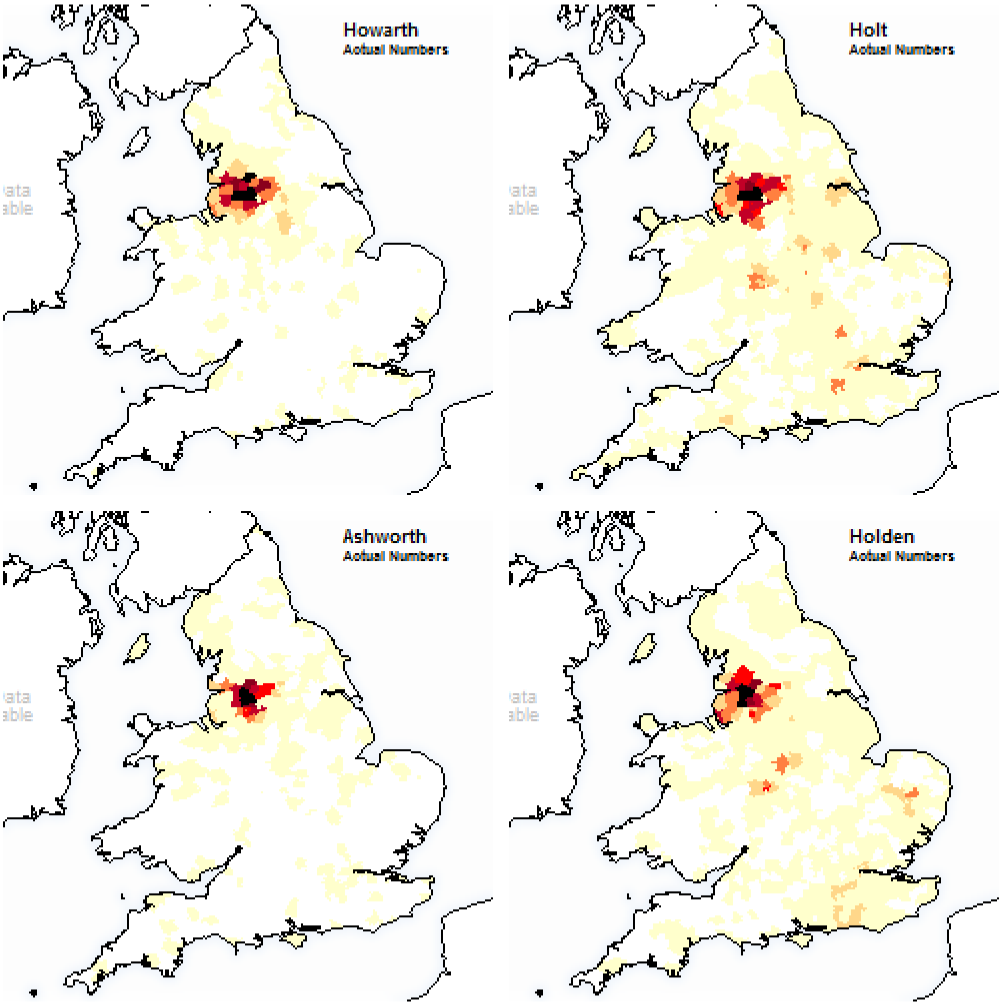



Figure 5: The four most populous single-ancestor contenders in Lancashire.

## Industrial Age population growth and the size of single-origin surname contenders

Maps are available showing the growth of the *total* population in each Hundred or Wapentake of the English counties, between 1761 and 1841.[22] From these, one might wonder whether, at that time of the Industrial Revolution, the exceptionally high growth rates in parts of Lancashire and West Yorkshire (between 500 to 750% growth) contributed to the high populations of the largest single-origin surname contenders there (Figures 4 and 5). Migration between Hundreds cannot be ruled out, however, as a significant factor in the population growths in specific Hundreds: new work opportunities could perhaps have attracted many to these particular areas of high growth in their overall populations.

[22] http://www.hpss.geog.cam.ac.uk/research/projects/occupations/englandwales1379-1911/figure2/figure2b.html

The leading single-origin surname contenders in Shropshire and Wiltshire (the two groups of four at the right side of Figure 6), for example, are strikingly less populous than in West Yorkshire and Lancashire (the two groups of four on the left). This might be compared with the general trends for overall population growths in the Industrial Age. We may note that particularly low growth of the general population took place in Shropshire (between 200 and 300% for the Hundred in that county with the most overall growth) and Wiltshire (between 50 and 100%) through 1761-1841. We might then proceed to wonder whether these low Industrial Age growths offer an explanation of the smaller sizes of the most populous surnames in these counties: the largest contenders on the right of Figure 6 are clearly much less populous, in line with lower growth rates in the counties, than is the case for the counties further to the right.

The bar-chart in Figure 6 shows the populations of the four largest single-origin contenders in respectively West Yorkshire, Lancashire, Cheshire, Staffordshire, Shropshire, and Wiltshire. This shows a general but not uniform fall, from left to right, for which the six counties are presented in the order of falling Industrial Age growth of their overall populations. Irrespective of whether they are single-origin or not, the four most populous surnames in Shropshire and Wiltshire, without omissions, are included in Figure 6.[23]

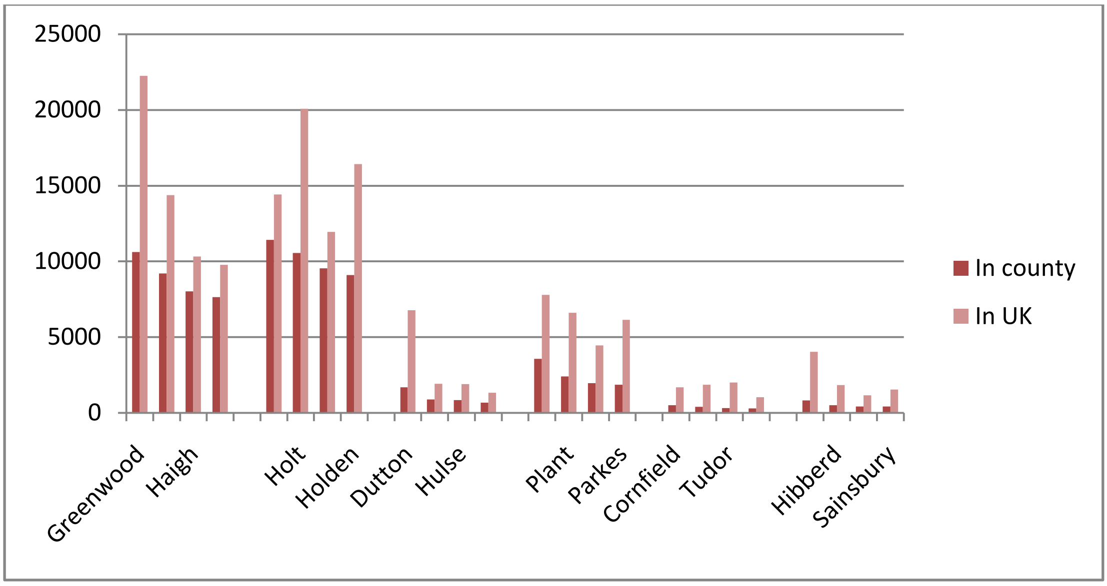



**Figure 6: Populations in 1881 of the four largest single-ancestor contenders in each of six counties: West Yorkshire; Lancashire; Cheshire; Staffordshire; Shropshire; and Wiltshire.**

[23] The top four for Shropshire are Cornfield (504/1,693), Gittins (411/1,855), Tudor (323/2,012) and Wellings (299/1,029). For Wiltshire, they are Hillier (821/4,038), Hibberd (498/1,839), Whatley (424/1,158) and Sainsbury (421/1,544).

From this irregular fall, it is clear that there are other factors, besides large Industrial Age growth, when considering the largest single-origin contenders in particular counties. For example, there was high growth of the overall population in *east* Cheshire (between 500 and 750%), comparable to that of the highest hundreds in West Yorkshire and Lancashire, and yet the most populous single-origin contenders in Cheshire (Figure 7) have significantly lower populations than those in Staffordshire (Figure 8) which had lower general growth, even in its south-eastern Hundred (300 to 400%). We will show later, for example, that overall growth throughout the whole of the county of Cheshire, rather than in specific Hundreds, was lower at these times than for all of Staffordshire. A persistent dilemma is also that possible effects of migration cannot be ignored, leaving the question of whether whole-county or Hundred growth rates are more appropriate, not to mention the possible importance of other conceivable effects on the growth of the apparently single-origin contender surnames.

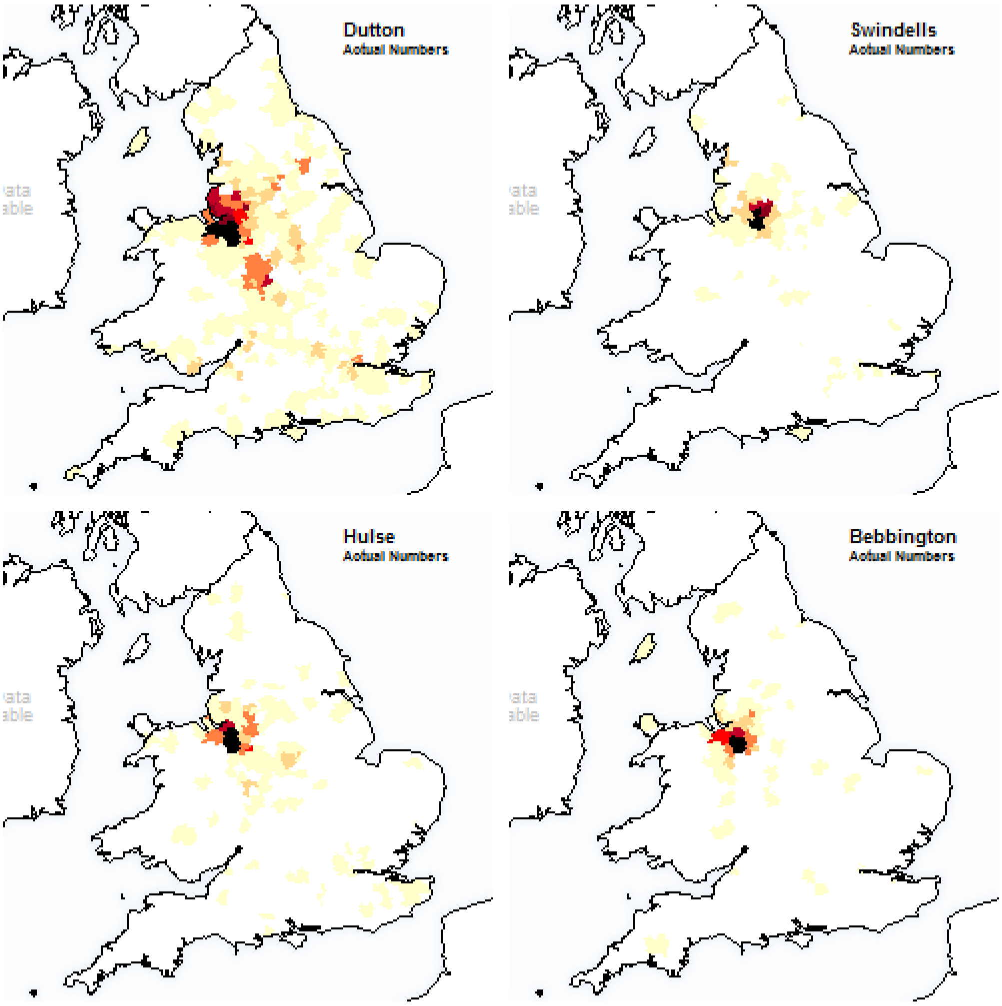



**Figure 7: The four most populous single-origin contenders in Cheshire.**

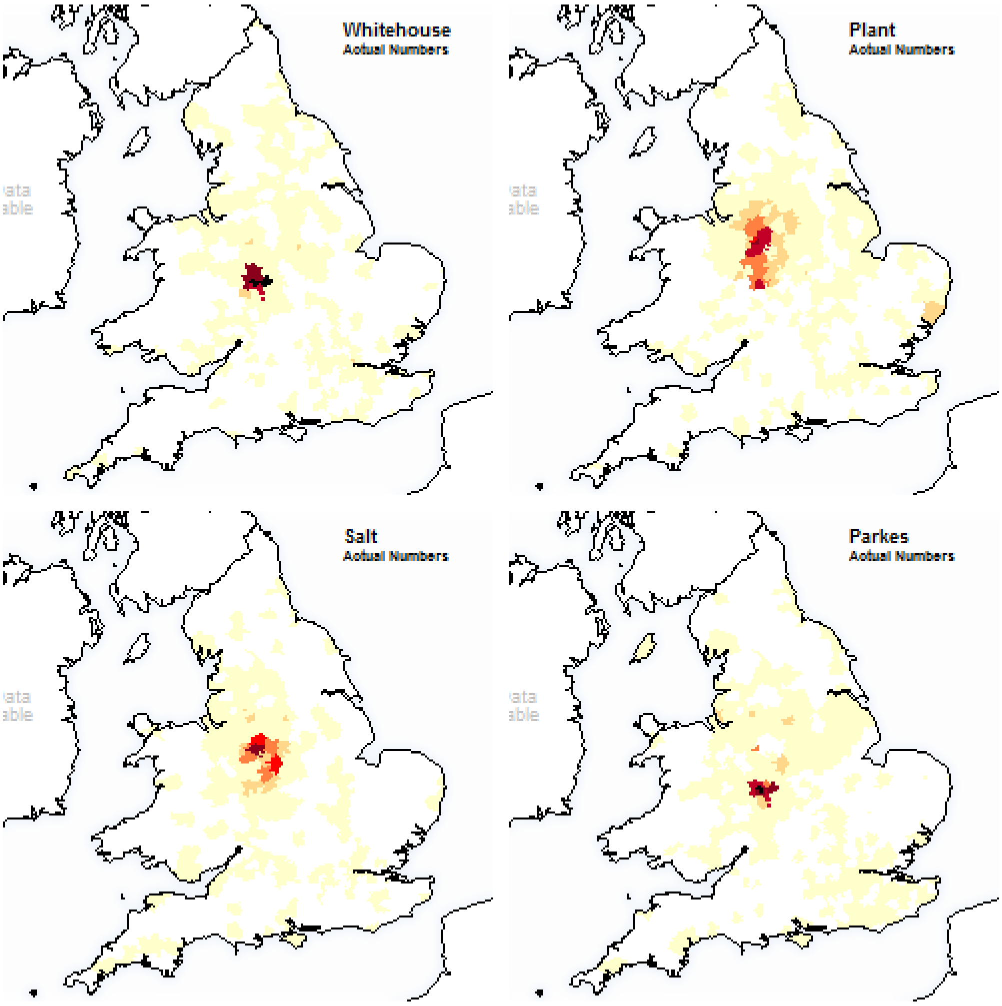



Figure 8: The four most populous single-source surname contenders in Staffordshire.

## Populous single-origin contenders in Cheshire and Staffordshire

The single-origin contenders that are included[24] in Figure 7, for Cheshire, are: Dutton (1,681/6,783); Swindells (885/1,915); Hulse (850/1,896); and, Bebbington (684/1,329). Of these, Dutton seems relatively widely spread and hence rather more doubtful as a single-family contender.

In Staffordshire, the most populous single-origin contenders have rather lower populations than those in West Yorkshire and Lancashire but considerably higher populations than those in Cheshire, Wiltshire and Shropshire. Figure 8 shows the distributions, in 1881, of the four most populous

[24] The only more populous surname, which is omitted for Cheshire, for being too widespread, is Maddock (728/2,746).

Staffordshire surnames that might be considered to contend as single-origin surnames. Their Staffordshire/UK populations, in 1881, are: Whitehouse, 3,576/7,787; Plant, 2,408/6,615; Salt, 1,963/4,417; and Parkes 1,857/6,143.[25] None of these is contained solely in Staffordshire, though their distributions show a concentration there. Plant apparently originated near the east Cheshire border with Staffordshire and migrated further southwards into Staffordshire, as will be described further later.

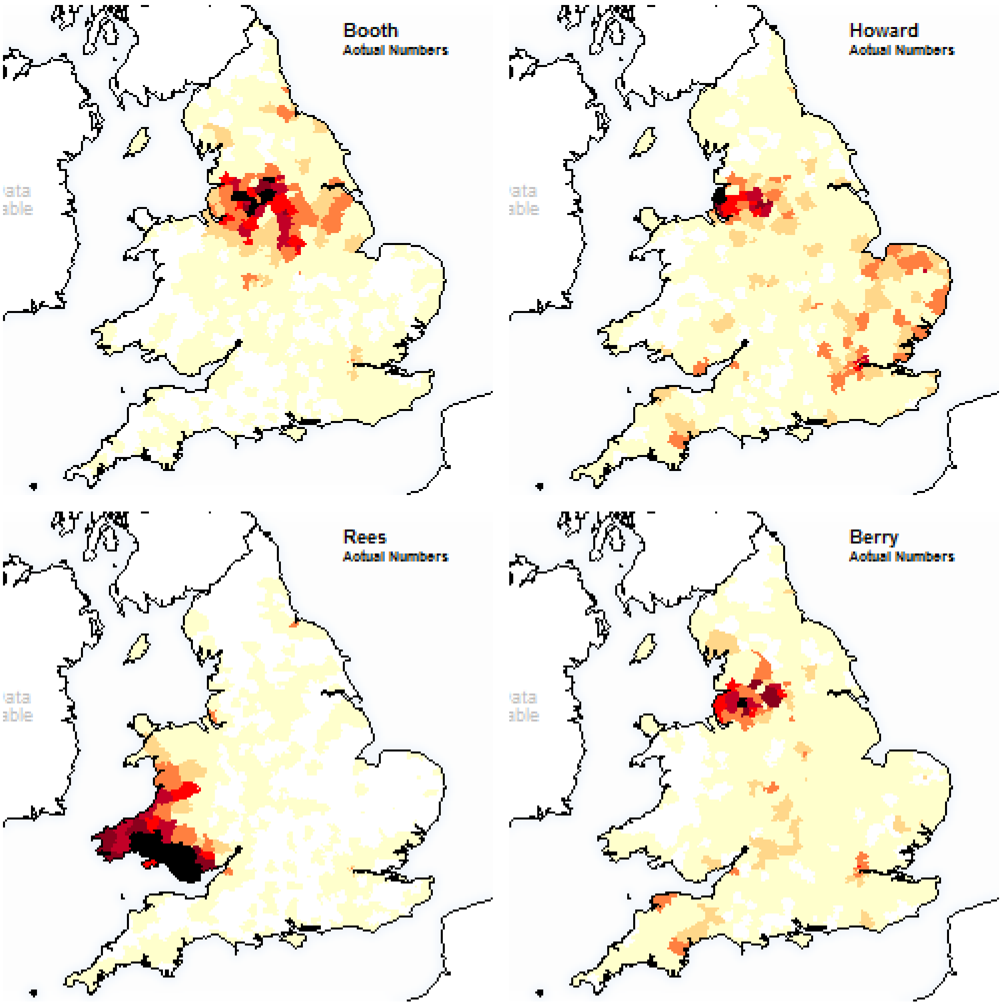



**Figure 9: Less likely, though the best of the feasible, single-origin contenders amongst the 145 most populous surnames in the UK in 1881.**

[25] Other populous surnames in Staffordshire, in 1881, are: Cartwright 2,493/11,406; and, Rowley 2,274/7,875. These are more widely spread, however, and they have accordingly been omitted from Figure 8.

### Some less-convincing single-origin contenders

In the preceding sections, the most populous single-origin contender is Greenwood (upper left in Figure 4) which is ranked 146th in the order of most populous surnames, with a UK population of 23,256 in 1881.

The following are more populous than Greenwood but they have rather less convincing distributions as single-origin contenders: Booth (ranked 106th, population 29,570); Howard (107th, 29,395); Rees (126th, 26,043); Berry (142nd, 23,775). Their distributions in 1881 are shown in Figure 9.

The 1881 surname distributions suggest that the first one hundred or so most populous surnames almost exclusively appear to be multi-origin, on the basis of their 1881 distributions. However, upon reaching the 150th surname, in the list of most populous surnames, a few single-origin contenders begin to appear to be *possibly* more debatable as single-origin contenders. Many of the most populous contenders appear to be in adjoining regions of West Yorkshire and Lancashire, with *perhaps* the most populous, on the basis of this subjective assessment, being Berry which is ranked 142nd and which has an 1881 UK population of 23,775. However, Berry might well be plural- or multi-origin instead, as has already been discussed in connection with an apparent meaning for this name.

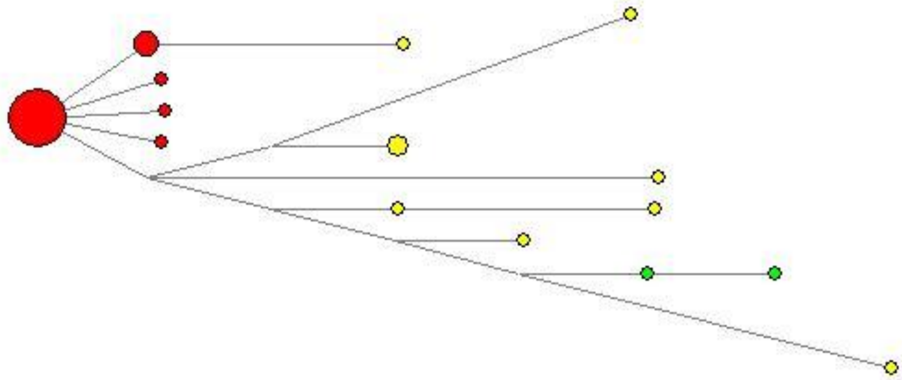
Taking another county, with moderately populous single-origin contenders, the most populous single-origin contender surname in Staffordshire appears to be Whitehouse, which is ranked 543rd with an 1881 UK population of 7,787, followed by Plant which is ranked 613th with a UK population of 6,615 in 1881. As will be discussed further below, there appears to be only a small plural-origin component contributing to the surviving population of the latter.

It is relevant to check the contention that these surnames could be single-origin, by means of appropriate DNA testing. The early DNA study for Sykes was, as has been shown, for the second most populous contender whose 1881 population is based mostly in West Yorkshire. In the next section, we shall consider in some detail the second most populous contender in Staffordshire: Plant.

## Considering a surname in more detail: Plant

As we have a wide range of detailed information for Plant, we shall consider this populous Staffordshire surname in some detail.

(a)



(b)

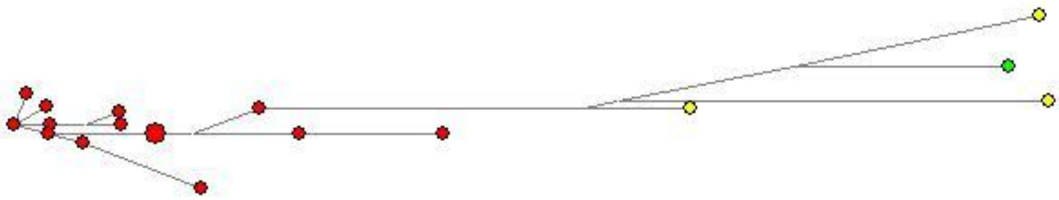


**Figure 10: DNA Network diagram for Plant: (a) 12-marker diagram; (b) 37-marker diagram.**

## The available DNA results

The 1881 map for Plant, upper right in Figure 8, shows it to be moderately widely distributed, with an apparent pattern of dispersal from a central location. The DNA network diagrams (Figure 10) indicate that there is a main cluster of matching Y-STR signatures, consistent with a single origin for many with this name. Some Plants have had 111 Y-STR markers measured and some as few as 10 but here we show just examples of the network diagrams that apply at the 12- and 37- marker levels. The Plants whose Y-STR results are included in Figure 10 volunteered independently, almost without exception. The exceptions are a pair who had shared ancestry in Ireland and another pair who both thought that their earliest known male-line ancestor had lived in Leicestershire, though one of these had only 10 markers measured and so does not appear in Figure 10 anyway. Some care has been taken in this DNA study to consider a random sample of the Plant surname.[26] Academic studies have used different procedures and conventional `snail' mail in trying to guarantee a random sample of a surname. On the other hand, the internet has been exploited for the Plant study.

The green circles in Figure 10(a) are associated with two participants with separate known ancestries, both in south Lincolnshire. Although no genealogical link has been found between them in the documentary evidence, 25-marker DNA testing confirms that they most probably have a shared male-line ancestor within surname times. Only one of them, who had 37 Y-STR markers measured, is included in Figure 10(b).

Otherwise, there is only one DNA cluster (red circles in Figure 10) yet found amongst the independent volunteers with the Plant surname. This is indicated by the large red circle and red outliers in Figure 10(a) and by the cluster of small red circles in 10(b). Their shared male-line ancestor might well date back to an early bearer of the Plant surname, perhaps located near the fourteenth-century "main" English Plant family homeland, or perhaps originating slightly earlier elsewhere, as is discussed later. There are three recognised methods of estimating the date of the shared male-line ancestor from the Y-DNA data and these all give a rough estimate of around the fourteenth century for the shared ancestor of those corresponding to the red circles.[27]

The large red circle in Figure 10(a) represents a number of exactly matching Y-STR signatures at the 12-marker level which is resolved by the extra markers in Figure 10(b) into a cluster of near-matching small red circles. The reason for non-exact matching of the red circles in Figure 10 is well understood: a small fraction of the Y-STR markers can be expected to have mutated, from an ancestral Y-signature, down the centuries of surname times. Two of the most distant outliers in 10(b), as well as two others less distant from the Plant modal Y-signature, have been investigated further at the 111 marker level; and, Deep Clade testing also confirms that they belong to the same surname family as those with fewer harmless mutations from the Plant modal haplotype. Those that match exactly in Figure 10(a) almost certainly belong to the main English Plant family and, when most of these and all but one of their red outliers have been upgraded to more markers, they have invariably remained matched to the main English Plant family.

There are those who do not even nearly DNA match with the main Y-STR cluster in Figure 10. For these non-matching yellow circles, one possibility is that some false-paternity events can be expected in the descent of any family. A false paternity event is often called, for example, a non-

[26] John S Plant (2009) *Surname Studies with Genetics*, DNA Section, Guild of One Name Studies (21 pages) esp. pp. 15-16. http://cogprints.org/6595

[27] http://www.plant-fhg.org.uk/dna_tmrca.html

paternity event (NPE) and it refers to any mechanism, such as infidelity or name change or adoption, by which subsequent Plants do not inherit a paternal Y-signature from the initial Plant-named male lineage since they have inherited the surname through a female link, perhaps because of wifely infidelity or an unmarried Plant mother passing on her own surname for example.

The green circles might correspond to an entirely separate origin for the Plant surname or, like each yellow circle, they might perhaps correspond to an NPE from the same surname origin as that for the red circles. Taking the latter case, in the 12-marker data (Appendix A) there are, at present, 23 matches (red circles in Figure 10(a)) and 11 non-matches (green and yellow circles) after discounting those who have had fewer than 12 markers measured and counting only one for the non-independent pairs of volunteers. Considering only independent volunteers, the *observed* non-matches comprise 32.4% of the dataset and this can be compared with *theoretical* expectation under the assumption that all the non-matches are NPEs. From such an assessment, it turns out that it is perfectly feasible, within the realms of theoretical expectation,[28] that surviving Plants descend from a single source for the name with all 32.4% of the mismatches being due to NPEs. We will discuss more fully later an alternative possibility that some of the mismatches might stem from other origins to the Plant surname, particularly for a dataset that applies more restrictively to just England.

None of the Plant volunteers has a Y-signature corresponding to the modal R1b1 signature of the UK general population. This R1b1 modal signature is the Y-DNA signature that is most frequent in the general population of Europe and North America. The available Plant results of those for whom only 9 comparable[29] markers have been measured are not far distant from this modal signature. However, the Plant results are distant from it when 25 or more markers are compared. Indeed, they have no known match with any tested non-Plant in the FT-DNA database at the higher levels of measured markers.[30] This supports the hypothesis that the red circles in Figure 10(a), or 10(b), correspond to descent from a shared ancestor who lived within the time span of surname history (say 800 years); they can be associated with a single ancestral source for the Y-signature of the main English Plant family. Apart from the green circles, there is no clear DNA evidence for another origin to the name, in contradistinction to many of the surnames studied by K&J.

---

[28] If the false paternity rate per generation is $p$, it can be *expected* that there will be $1-(1-p)^n$ non-matches after $n$ generations. If we assume 23 generations since the origins of the surname, we get a value of p=0.016, to wit an NPE rate of 1.6% per generation. This is within the limits of a "reasonable estimate" of 2% as assumed for a "typical" surname and used by K&J in their computations.

[29] All had at least 10 markers measured but only 9 of these could be universally compared.

[30] There is one with a different surname who matches at the 37-marker level but subsequent genealogical research has shown him to be a lost close relative, adopted when young, of another Plant who has been Y-DNA tested.

The modal Y-STR signature for Plant, at a 38-marker[31] and higher levels, is given elsewhere.[32] Plant volunteers in the main DNA cluster have been found to belong to the R-P312+ sub-clade R-DF27+(xZ196,xZ225).[33] We will discuss later below how this provides information about the deep ancestry of the main matching Plant family.

## Is the Plant surname single origin?

Before the advent of DNA testing, Plant was thought to be multi-origin because it was populous and moderately widespread.[34] At that time, the term "single-origin" was usually reserved for relatively rare surnames that were almost entirely in a small local geographical area.[35] There was also a contention that Plant meant a `gardener' and it was accordingly presumed that there were many origins corresponding with many different gardeners. However, this contention has now been thrown into substantial doubt. As early DNA evidence became available, it offered conflicting evidence that indicated that Plant (top right in Figure 8), perhaps like Sykes (top right in Figure 4), might be predominantly "single origin".[36] By this, it is meant that the DNA-tested *living* Plants can mostly be ascribed to male-line decent from a single male ancestor, who lived within the past 800 years or thereabouts.

In fact, a name such as Plant might have had a few origins in medieval times though the DNA evidence for living Plants displays a distinct single main cluster (Figure 10). This DNA clustering more strictly indicates that one family has survived more prolifically than others down to the living bearers of the surname who have been DNA tested.

---

[31] The number 38 comes from 37 markers measured by Family Tree DNA (FT-DNA) and an additional marker, not included by FT-DNA, which was measured besides 9 others by Oxford Ancestors. A modal signature has also been determined at the 68 marker level and, with less certainty, at the 112 marker level.

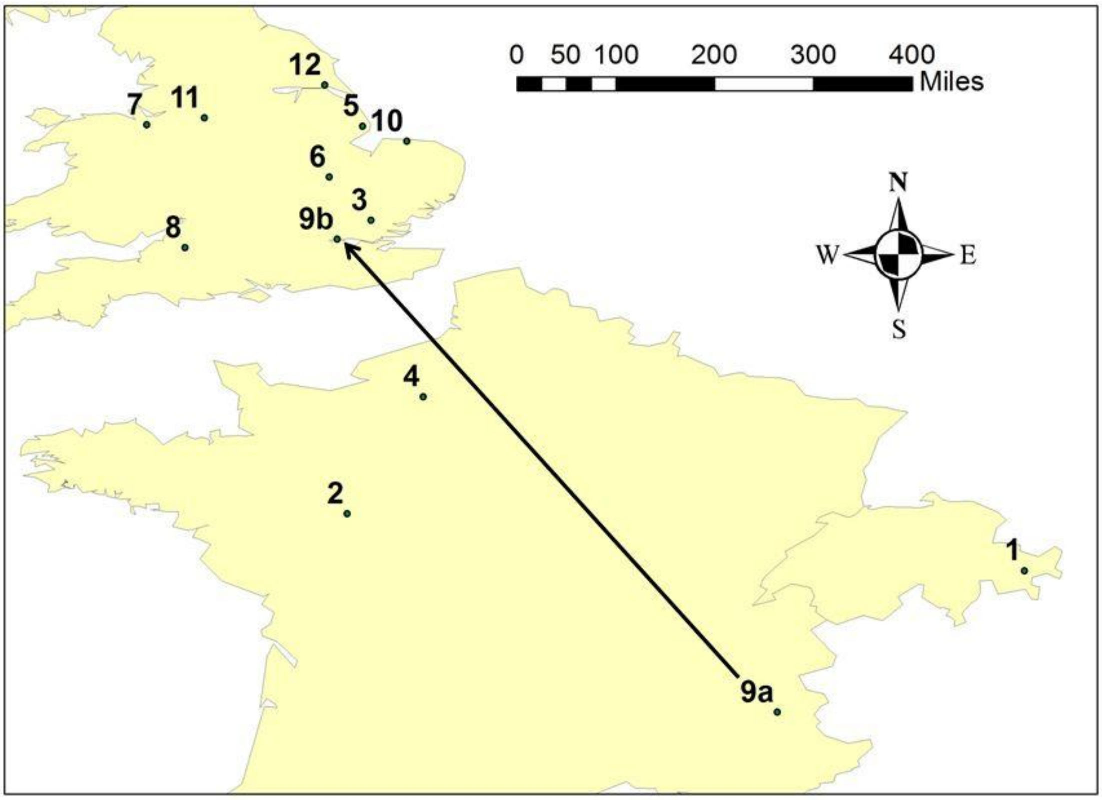
[32] Fuller details of the DNA results for Plant are given at http://www.plant-fhg.org.uk/dna.html .

[33] Here, R-DF27+(xZ196,xZ225) means that the DF27 has tested positive but its sub-clades Z196 and Z225 (and, in fact, some of their sub-clades) have tested negative.

[34] The nineteenth Earl lecture, delivered at Keele University, 6 November 1997; published as David Hey (1998) *The distinctive surnames of Staffordshire*, Staffordshire Studies, 10, pp. 1-28, esp. 14.

[35] David Hey, *Family Names and Family History* (Cambridge University Press, 2000).

[36] John S Plant (2005) *Modern Methods and a Controversial Suname: Plant*, Nomina, 28, pp. 115-133.




**1** *1139-1798* Seat of the noble *Planta* family in the Upper Engadine

**2** *1202* Lands at Chinon and Loudun of Emeric *de la Planta* alias *de Plant'*

**3** *1262* First known evidence of the name in England; spelled *Plaunte*

**4** *1273* Three Rouen merchants called *de la Plaunt* and *Plaunt*

**5** *1279* At Burgh-le-Marsh near Bolingbroke, the name *Plante* is indicated to have been hereditary for 3 generations

**6** *1282* The name form *de Plantes* in Huntingdonshire

**7** *1301* First evidence of the *Plant* name local to the subsequent main homeland of the surname

**8** *ca.1280-ca.1360* Records of *Plonte* name at Bath, explicitly hereditary by 1328

**9** *1350* London priest Henry *Plante* of Risole: **9a** is Risoul; **9b** is London

**10** 1352 James *Plant* carried away goods from recently lost Warren lands in Norfolk

**11** 1360 onwards Several records of *Plonte* or *Plont* in the subsequent main *Plant* homeland

**12** *1379* A gardener called *Plant*

**Figure 11: Some medieval records for Plant-like names. See http://www.plant-fhg.org.uk/origins.html#13c for a fuller list and details.**

The DNA results for the living Plants might not relate to all of the first Plants. There might have been some unrelated Plant families that died out or proliferated little so as not to show up as separate distinct clusters in the DNA results for the living Plants of Figure 10. We shall accordingly call Plant an "*effectively* single origin" surname, though it might well have been plural-origin in medieval times. The Plants with mismatching y-signatures might descend either from random non-Plant men fathering NPEs within the single Plant family or some might indeed descend from different origins of the Plant surname. The two properties that characterise an effectively single origin surname are that the DNA results are numerically consistent with the non-matches being NPEs, and that the non-matches to the dominant cluster do not display any structure that would indicate a second familial relationship or more. We shall return to this question in the subsequent computer modeling of the surname.

Figure 11 shows that medieval records for the Plant name were spread throughout France and England. The medieval records indicate that there were early Plants in such classes as the clergy as well as freeholders and merchants and a bailiff.[37] Early descent from one such Plant family could have given rise to a relatively widespread "effectively single origin" surname. However, though there were Plants with the means for distant travel (e.g. medieval Plants at 1, 2, 4, 7, 8, 9, and 10 in Figure 11), this does not constitute convincing evidence that all early instances of the name in the medieval records belonged to the same Plant family.

Some early instances of the Plant name could have been by-names that were never inherited; and, even initially inherited lines can be expected to have died out, entirely or largely, to a dramatic extent, as our computer simulations show. By 1400, there is explicit evidence that the English Plant surname was, or had been, hereditary in Lincolnshire (item 5 in Figure 11), at Bath (item 8) and around the northern border of Staffordshire with Cheshire (item 11). This might be taken to suggest three different origins for the surname in England. On the other hand, the social status of these Plants was sufficient that we might venture to postulate that they *could have* migrated from a single origin. Those with the hereditary surname Plant (Plonte) around Bath might have died out (or perhaps become bearers of a morphed version of the surname, such as Plenty); those in south Lincolnshire may or may not have survived as more modern south Lincolnshire Plants; and, those on the Cheshire-Staffordshire border apparently became the main English Plant family. Our preferred presumption is that this main Plant family was not already more widely spread before it migrated to its main homeland, though we shall also consider an hypothesis that it might have been.

---

[37] John S Plant and Richard E Plant (April-June 2012) *The Plant Controversy,* Journal of One-Names Studies, Vol. 11, Issue 2, pp.12-13.

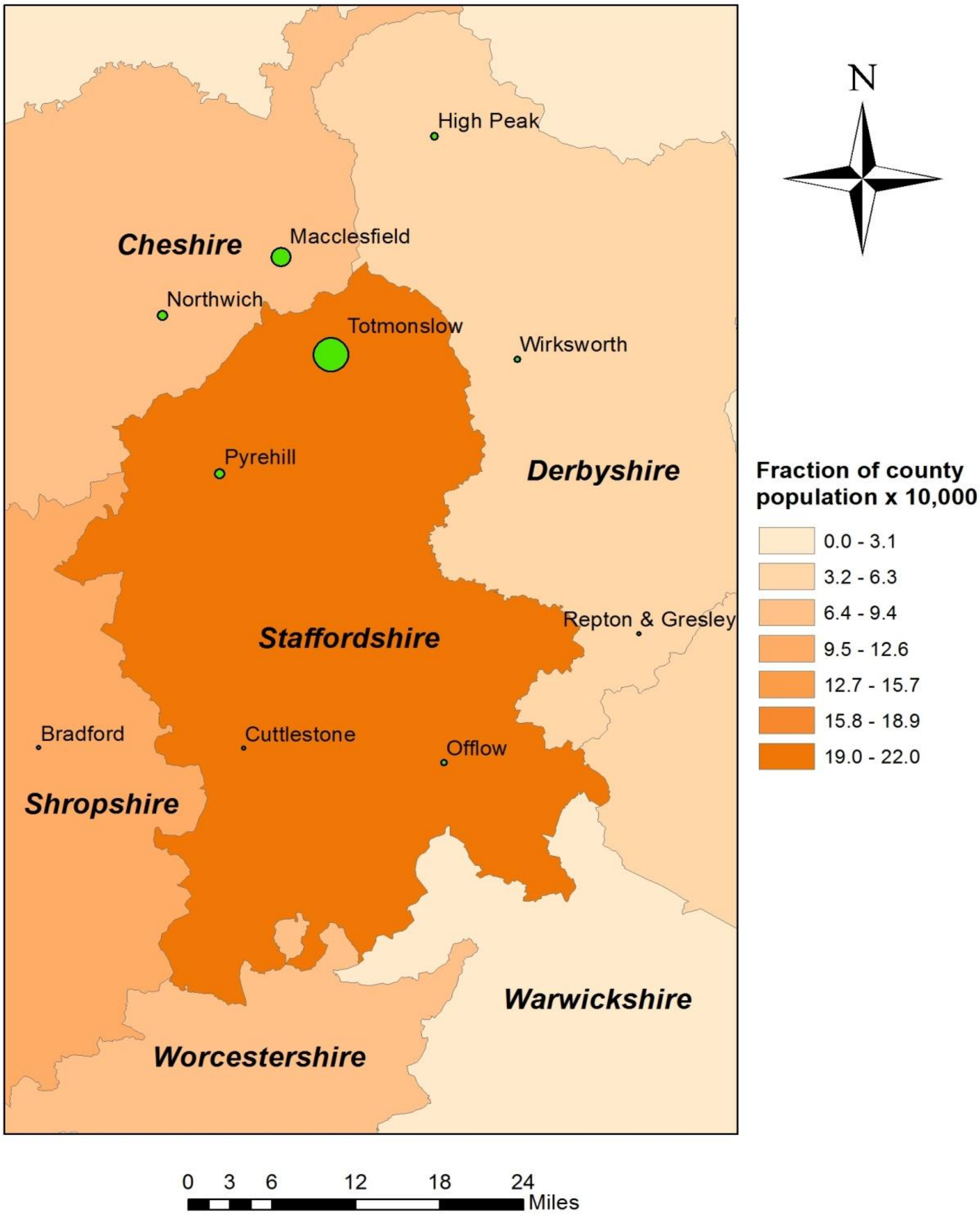



**Figure 12: A cluster of Plant Hearth-Tax households (1662-89) is represented by the green circles. The graded brown represents the number of Plants per 10,000 in each county, in 1881.**

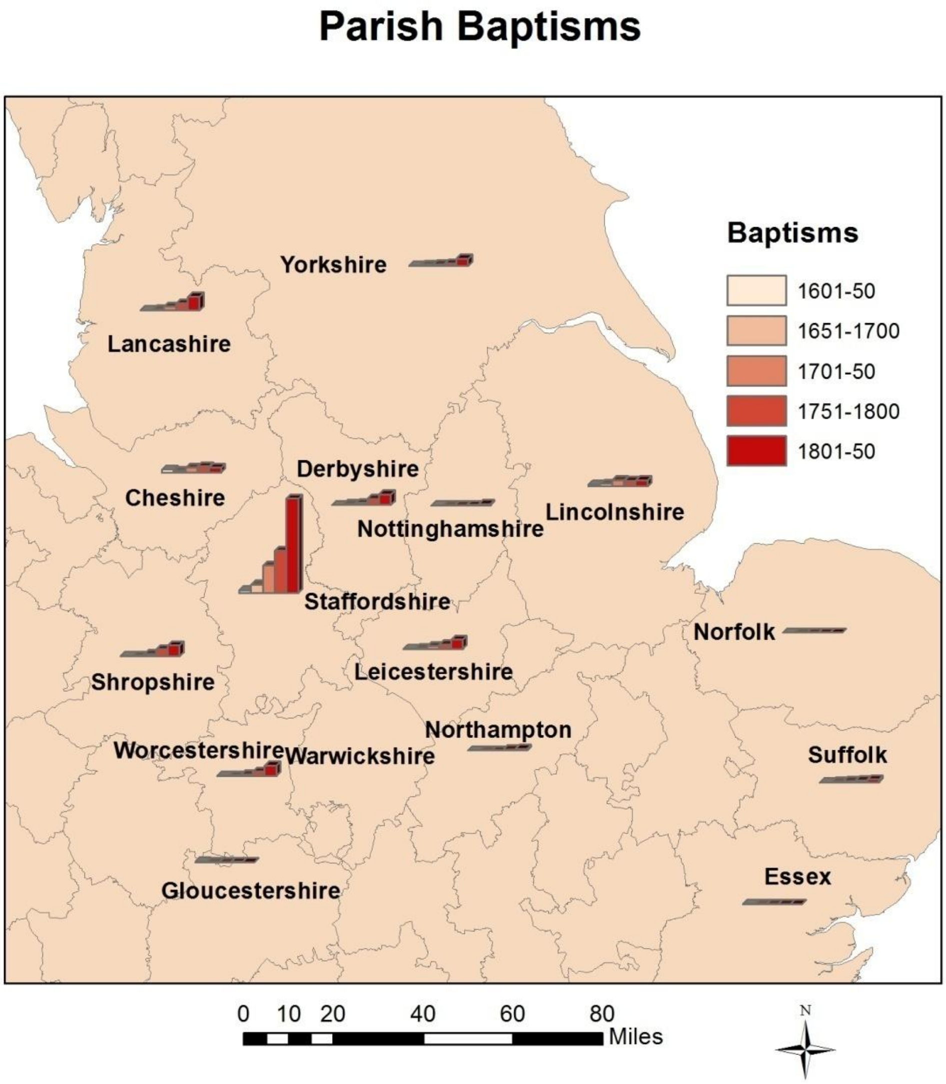



Figure 13: Distribution of Plant baptism records throughout several English counties (1601-1850).

The medieval Plant records at item 11 in Figure 11 correspond in location to a cluster of seventeenth-century Hearth Tax records; these are found particularly in Totmonslow and Macclesfield Hundreds in the map shown in Figure 12. Hearth Tax records are not readily available for the whole of England and those shown in Figure 12 are only for Hundreds around the main Plant homeland. We refer to the area of the larger green circles in Figure 12 as the main homeland of the main English Plant family. The “item 11” location of the medieval records of Figure 11 corresponds closely with the large green circles as well as the peak of the 1881 distribution of Plants in Figure 8.

The other marked points in Figure 11 show a much wider distribution of early instances of Plant-like names - these can be regarded as separate medieval origins to the name that died out though it remains open to speculation that they could conceivably be early locations for a mobile family.

It is not inconceivable that early Plants travelled far and that their name dated back to very early times. It has even been claimed[38] that the name Planta (item 1 in Figure 11) dates back intact to a Roman inscription for Julius Planta dated 46AD at Trento, near the Engadine seat of this noble family (1139-1798). This is very contentious, even though it is somewhat more credible that the late medieval name Planta *might have* morphed to Plante and then Plant.

Certainly, there is evidence that the Plant name was quite widely spread by early modern times. The mapped bar-charts in Figure 13 display the growths in listed Plant baptisms in the IGI,[39] from 1601 to 1800, in various counties. This growth is most notable in Staffordshire. However, fairly consistently, about half of the Plants are in the main homeland of Cheshire and Staffordshire and half thinly spread elsewhere, though it should be borne in mind that there are shortcomings with this IGI data and too precise an interpretation of Figure 13 is inappropriate.

The mapped bar-charts in Figure 13 indicate that there were Plants in several counties by the beginning of the seventeenth century. Their numbers grew throughout the eighteenth century, especially in Staffordshire, leading on to the distribution shown in Figures 8 (and by the brown shading in Figure 12) by the times of the 1881 Census. In Figure 8, the 1881 Census data is aggregated into Poor Law Unions whereas a more broadly smoothed picture is presented in Figure 12, where the 1881 Plant population data (brown shading) is aggregated into whole counties and scaled to represent Plants per 10,000 of the total county population. The bar-charts in Figure 13 indicate the different numbers of Plants in different counties, using data which are not reliably complete but which, with this word of caution, add more of an historical and geographical perspective to the growth in the numbers of Plants.

[38] G.R.de Beer (1952) Notes and Records of the Royal Society of London, p.8.

[39] These data have been taken from the 1984 microfiche of the International Genealogical Index.

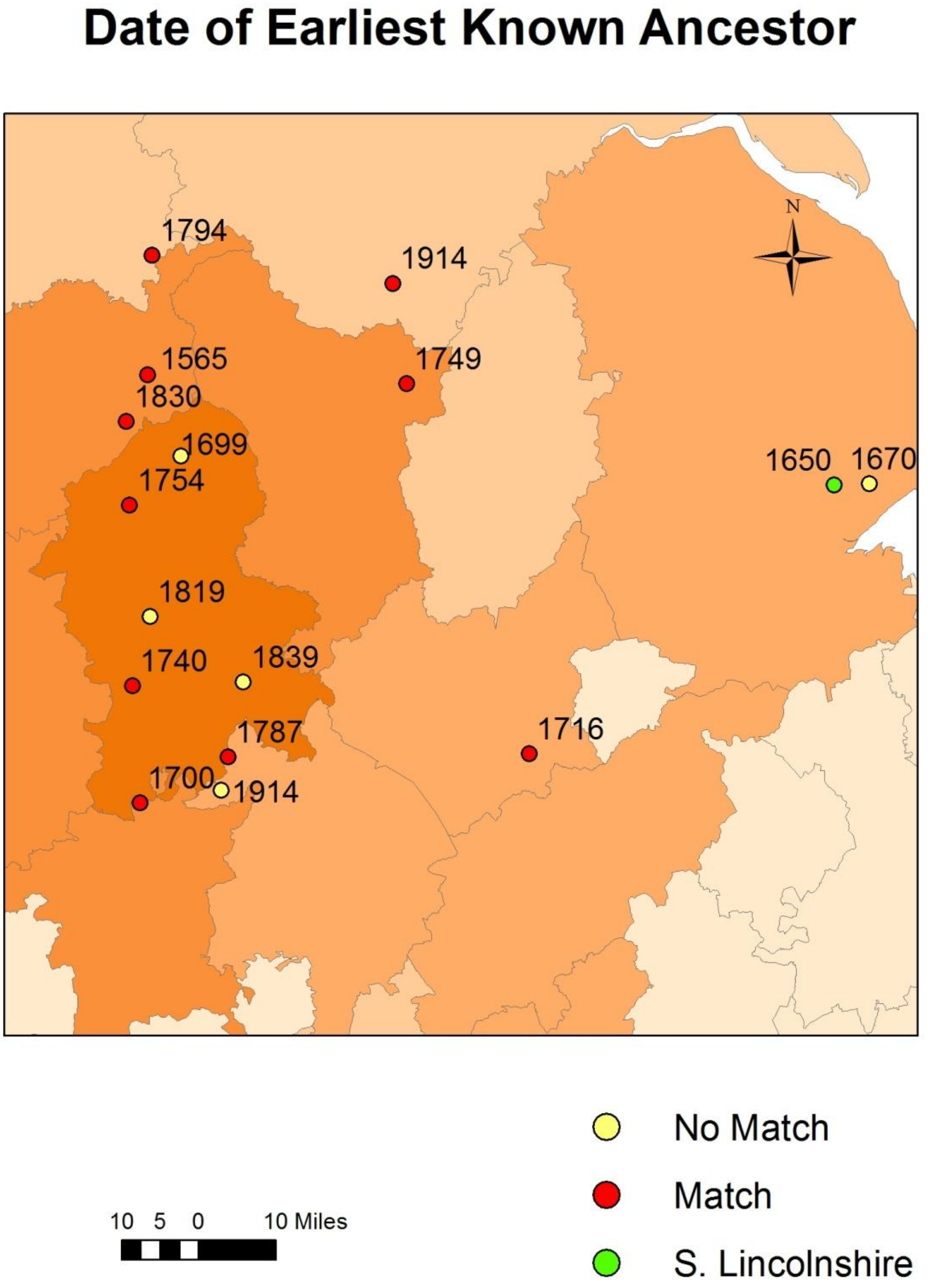



Figure 14: Distribution of DNA matching and mismatching ancestral Plants. Where two matching Plants have been found in the same local region only one circle is here included. The labels are the dates of the earliest known Plant ancestor at each location for those who have been DNA tested.

## Relating the DNA results to the geographical distribution of the Plant surname

Figure 14 shows the distribution of earliest known ancestors for the DNA matches and mismatches of some tested Plants. As in Figure 12, this is superimposed on brown shading that depicts the smoothed distribution of the Plant population, relative to county totals, by 1881.

The distributed circles in Figure 14 correspond to independent volunteers who represent a roughly random (not uniform) sample of the Plants. In fact, three different Plants volunteered independently who each thought that their ancestry was in NE Derbyshire. It might just be a random coincidence that all three volunteered independently thinking that they came from the same place. However, only one red circle is shown in Figure 14 for these matching Plants. Also, only one red circle is shown for a pair in south in Leicestershire, who volunteered together and so cannot be considered to be individually independent. The green circle corresponds to a pair whose separate ancestral lines coexisted in south Lincolnshire; they volunteered independently and, though they did not match with the main English Plant family (red circles), they matched one another at the 25-marker level. Whereas the red circles represent DNA matches, the yellow circles correspond to volunteers who DNA match neither the main paternal Plant family (red circles) nor other yellow circles nor the green one.

The dated red circles in Figure 14 show that there was a widely-spread single family of DNA matching Plants by around the eighteenth century. The circles are labeled with the dates of the earliest known male-line ancestors of these tested Plants and they show that broadly, by the eighteenth century, the matching single Plant family extended through and around the peripheries of the county of Staffordshire. Indeed, there were other matching Plants in Ireland (ancestors traced back to 1808) and the USA (back to 1655). The full detail of when and how this single family of matching Plants ramified widely, before the eighteenth century, can only be guessed, with the help of the mapped geographical distributions of *all* Plants (e.g. Figure 13).

Attempts to trace back the earliest known ancestors, of those Plants who have been tested, to earlier dates than the ones already included in Figure 14 might in time play a role; but, the ease and reliability of such a genealogical extension, back towards early times, should not be exaggerated for this populous surname.

## Possible relevance of various meanings of *plant*

There are four possible meanings of *plant* that might relate to the origins and development of the Plant surname. Two of these hypotheses relate to the earliest origins of the name, which as Figure 11 indicates could have been either very far from (e.g. item 1 in Figure 11), or near to (e.g. item 7), the main homeland of the surviving Plant surname (item 11 in Figure 11). Two other semantic hypotheses relate, in turn, to discussions as to whether either a metonymic extension of the word's meaning,  or a newly developed  industrial meaning for *plant*, might be associated with a plural-origin contribution to the surviving Plant population.

First, considering the possibility of a distant initial origin to the name, the form *de la Planta* (item 2 in Figure 11) might mean from La Planta region, of the Alps for example. This might refer, for instance, to the Engadine which means, in Romansh, `garden source of the River Inn' (item 1 in Figure 11). The name form *de la Planta* evidently developed into *de Plant'*, *de la Plaunt*, and *de Plantes* in England (items 2, 4 and 6 in Figure 11) though it has alternatively been suggested that *de Plantes* might mean from Le Plantis in Normandy . The Avignon Popes (1309-76) tended to favour appointments from their local region, which might explain why Henry Plante of Risole, in the French Alps, was awarded the stipend of a priest in London (item 9 in Figure 11). To this extent, the semantics and historical evidence suggest one particular hypothesis that lesser members of the noble Planta family *might have* supplied a "single origin" to at least some of those surviving with the English Plant surname.

As another hypothesis, the by-name Plant has been said to mean a `gardener' (cf. item 12 of Figure 11). This relates only by theoretical metonymy to the accepted medieval meanings of the word *plant*. Such a meaning, of course, does not appear in language dictionaries. As we shall show later below, it is not necessary to assume that Plant is a multi-origin occupational name, derived from many different gardeners, in order to explain the large population of the Plant surname. At Hull (item 12 in Figure 11) there might have been a connection with gardening made to the Plant name for the *gardiner* there called Plant in 1379 but other Plant records give contrary evidence. It is not clear that this gardener meaning was an origin to the hereditary surname of the dominant Plant family or indeed that it contributed at all to the meaning for any of the surviving population of Plants. It can be noted, for example, that recent evidence indicates that the Plant name (with the dialect spelling Plont) existed in the main Plant homeland (item 11 in Figure 11) at least two decades before the isolated evidence for a gardener at Hull (item 12).

Nearer to the main Plant homeland, we can invoke the Welsh meaning `children' of *plant*. Before the meaning `gardener', popular in recent Surname Dictionaries, the Plant name was said to mean `young offspring'. Such a meaning is compatible with evidence of a Welsh influence in the main Plant homeland; this evidence can be supplemented by that of the proximity of an early Plant in Wales (Item 7 in Figure 11). A putative single male-line ancestor might have fathered a family or `clan' before the Plant surname stabilised. In other words, the fore-fathers of the surviving Plant surname might have come largely from a pre-existing stock of related men who inhabited a local geographical area, though not necessarily with the Welsh meaning to their name. Relaxing the "single origin" contention, these could have supplied a few origins to Plant as a hereditary surname, some of whom closely DNA matched and some of whose matching lines survived down to form the bulk of the now living Plants.

Various meanings for Plant can be set in the context of the word's earliest etymology. According to Claiborne,[40] *plant* derives from the Indo-European root *plat-* meaning `flat'. From there it passed to the Latin *planta* meaning `sole of foot', and from there to the later meaning `seedling', which was pressed into the ground with the foot. The word passed into Gaelic where the initial *p* underwent a sound shift to *c*, resulting in *clan*, meaning the offshoots of a family. Others have pointed to ancient beliefs in the emergence of human life from the land with a lame foot,[41] suggesting sole (of foot) as an ontological foundation to man's existence or soul. Certainly, the word *planta* means `to breed'[42] or `to beget children' in Welsh in which *plant* means `children' and, in contemporary belief, this can be related to the `planting' of soul from both God and the human father.

In contemporary Middle English, the word has a rather wide range of meanings, some with religious connotations, as outlined in the Middle English Dictionary and the Oxford English Dictionary. A rather complex example of the Middle English usage of the word *plant* can be illustrated with an extract from William Langland's late fourteenth-century poem *Piers Plowman*. This poem states that `love is the plant of *pes* [peace or peas]'[43] which could well allude to Jesus as the vine or divine root; the poem later adds `this *tre* [tree or trinity] has true-love ... this is a proper plant that brings forth folk of all nations'.[44] To help explain this, it can be noted that the late medieval meanings of *plant* conflate several modern `biological' concepts. The word *plant* meant a shoot or an offshoot or a vine or a tree; and the poem, with its medieval philosophy, conflated these now-diverged meanings with the generation of offspring through the soul. Contemporary belief in the human soul involved the generative power of the vegetative seed as well as the divine intervention of God's planted Word.[45]

We can consider another developing meaning of *plant* as well as the possibility of a late conflation with the supposed root *plat*. Either might be related to a small multi-origin influx of population into a pre-existing Plant surname. Relatively late aliases and name changes are not unknown and they have been claimed to be more common than is generally supposed.[46] Two such possibilities are as follows.

---

[40] Robert Claiborne, *The Roots of English: A Handbook of Word Origins* (1989).

[41] C Levi-Strauss, *The Structural Study of Myth*, Structural Anthropology, translated from French by C Jacobson and B G Schoepf (New York, 1963), pp. 206-231.

[42] Also, Samuel Johnson's 18th century English Dictionary gives the meaning `to procreate' for *plant*.

[43] C-text version of *Piers Plowman*, Passus II, l. 149; in the Huntingdon manuscript HM 143, as listed by E Salter and D Pearsall, *Piers Plowman* (London, 1967), p. 73.

[44] Huntingdon manuscript HM 143, *Piers Plowman*, Passus XIX, ll. 9, 25-26, 101-102.

[45] http://www.plant-fhg.org.uk/soul.html

[46] George Redmonds, *Surnames and Genealogy: A New Approach* (Boston, 1997).

- The surname Platt was concentrated just to the north of the main Plant homeland and it is not inconceivable that there might have been some lexical confusion of Platt with Plant.
- More generally, there might have been a multi-source influx from any other surname that was not yet fully fixed and, in particular, this can be related to a tenuously supposed attraction of a relatively late semantic development of the word *plant*.

The Oxford English Dictionary[47] records a new eighteenth-century meaning for *plant*, which is the tools and equipment required to generate an industrial process. There is general controversy about how this meaning came about, though one might point to a possibility that the generation of an industrial product could have been metaphorically based on the idea of the generation of an offshoot or child or young animal – the verb *to plant* had meanings `to found' or `to establish' from early times, according to the Oxford English Dictionary, and these can be related on to industrial-scale generation as evidenced in the same Dictionary for later times. We might wonder if such a meaning could have appealed to industrial-minded farmers or tradesmen around the Potteries of North Staffordshire, for example, and could perhaps have been attractive enough to induce them to change their family name. Such name changes to Plant so far lack confirmation however, as does the precise way in which the industrial meaning of the word *plant* developed.

It might be relevant that industrial associations perhaps existed early in this region with a sense `to set' of *to plant*. For example, by the seventeenth century, there was engineering in the nearby lead mines of Derbyshire, with `Engines, Pumps, Forces there set and planted'.[48] As further speculation on the possibility of an early industrial sense to the word *plant*, we may note that French was pervasive in Middle English and that the French meaning of *plante* is a `bed or planting place'; to this we may add that the first known location in the main Plant homeland is for Thomas Plontt at a vaccary (cattle breeding and rearing station) such that *Plant* might mean`from the planting or breeding enclosure', possibly leading on to an industrial sense in industrial-scale farming as well as forming a possible explanation of the initial meaning of the surviving Plant surname.

The supposition of a late multi-origin influx into the Plant surname is not necessary to explain the DNA evidence or the prolific growth of the Plant population however. The possible causes of such growth will be investigated more fully in the Modelling section of this paper below.

## Did the Plant family arrive in its main homeland from elsewhere?

Recently extended searches of the Macclesfield Court Rolls of the main Plant homeland (item 11 in Figure 11) have revealed evidence that is relevant to the `offspring' and `gardener' meanings, as well as to a hypothesis of how Plants might have arrived in their main homeland.

Some key records amongst the earliest yet found for Plant at Macclesfield[49] can be summarised as follows:

---

[47] OED, *plant* n(1) 6a.

[48] Nellie Kirkham, *Derbyshire Lead Mining through the Centuries* (Truro, 1968), p. 88.

[49] http://www.plant-fhg.org.uk/EarlyPlontsMacclesfield.pdf

- 1360 Rand Plont(t) and Willo Plont(t) fined for stray animals in the Forrest
- 1363 Thom Plont(t) indicted; then in 1365, as an outlaw, committed to prison pending paryment of 20s (i.e. £1)
- 1370s Seven Plonts are mentioned in 35 separate entries for stray animals and pannage of pigs
- 1374 Thom Plontt had failed to pay the fine for pasturing a bullock in the Black Prince's vaccary
- 1383 Ranulph Plont renting lands in Rainow formerly belonging to John Walshe
- 1401 Richard Plont is guarantor for John Togard and Nicholas le Gardiner; Ranulph Plont is guarantor for Nicholas le Gardiner

To these can be added a few records for just across the county border into Staffordshire:

- 1379 Thomas Plonte aided and abetted a beheading on behalf of the Abbot of Dieulacress Abbey but obtained the King's pardon in 1382
- 1397 John and Richard Plont sued for trespassing with herds of cows just across the River Dane from the Black Prince's erstwhile vaccary
- 1406 Edward Plont granted *two mess' one croft* by the Abbot of Dieulacress

Though these records do not generally give clear familial relationships, it can be noted that there are two relevant associations of the Plant name with other names. The surname *Walshe* (item dated 1383 above) is believed to mean `from Wales' and the Welsh meaning of *plant* is `children', not gardener, for which the name *le Gardiner* is here in evidence (item dated 1401 above). This does not prove a semantic association of Plant to either `offspring' or `gardener' though it does suggest that both of these meanings were in play in a direct social association with the Plants. It is not impossible that the name, here, initially meant `offspring', particularly in the context of the name *Walshe*, and then developed to mean `gardener' in more of an assimilated English context, if it did not simply mean `from the planting enclosure' in connection with the Black Prince's vaccary at Midgley.

Less directly, there is also an indication that is not incompatible with a lexical confusion between Platt and Plant. There is a 1357 record at Adlington, five miles north of Macclesfield town, for a Richard *fil Plot*.

The name of the main Plant family could have originated in Macclesfield Forest in east Cheshire. Alternatively, however, it might have arrived from elsewhere. For example, it is possible that it arrived with the import of cattle from Wales (item 7 in Figure 11) to the Black Prince's vaccary for fattening[50] at Midgley on the Cheshire-Staffordshire county boundary (items dated 1374 and 1397 in the above lists), at the heart of the main Plant homeland. A subsequent connection with Dieulacres Abbey in Staffordshire (items dated 1379 and 1406) was perhaps initially associated with the

[50] Clarice Stella Davies (1976) *A History of Macclesfield*, 12.

Cheshire outlaw Thomas Plontt who apparently sought sanctuary across the county boundary to the south. This might have aided a prolific spread of the Plant family into the considerable land holdings of Dieulacres Abbey in Staffordshire. The 1881 distribution (Figure 8) indicates the Plant name's distribution had subsequently elongated progressively further southwards, reaching as far as the region of the large city of Birmingham just to the south of Staffordshire.

Though far more contentious, another hypothesis has been that the medieval name was displaced to the medieval Cheshire-Staffordshire border from Warren lands elsewhere.[51] The last Earl Warren died in 1347 and, further to a 1342 marriage to the Poynton heiress, his illegitimate descent inherited Poynton in Macclesfield Hundred in 1370. Earlier, the Earl Warren had briefly held the High Peak, adjacent to the main subsequent Plant homeland (position 11 in Figure 11). Displacement to here could have stemmed from the Lancastrians taking over Warren lands after the last Earl Warren's death in 1347. One piece of supporting evidence is that, in 1352, James Plant is mentioned for carrying away goods from a recently lost Warren Hundred in north Norfolk (item 10 in Figure 11). Taking an association with Warren lands back further, the Plants at position 8 in Figure 11 were near disputed land during the early fourteenth-century feud between the Earls Warren and Lancaster. Also, position 7 was near the Broomfield and Yale lands of the Earl Warren following the thirteenth-century Welsh Wars. The further back one goes the more tenuous the evidence becomes but one might also note, for example, that the first known hereditary evidence for the Plant surname in England (item 5 of Figure 11) was also near Warren lands. Moreover, in the mid thirteenth century, these lands were under the Wardship of Peter of Savoy, who came from near position 1 in Figure 11. This tenuous possibility of an ultimate single origin, feeding into England, could be supposed to have led on to a subsequent distribution of Plants with roles of modest rank in connection with the Warrens' lands. However, though the Plants were repeatedly found near such lands, these coincidences could have been just fortuitous and they cannot be regarded in any way as `proof' that the main Plant family's name initially meant `from the Planta region of the Alps', though early forms of the name might have originated with such a meaning.

It might also be relevant to add a few words about the current thinking, though still developing, for the DNA haplogroup of the main Plant family. Some archaeologists imagine that the parent *ancient* familial grouping, to wit the haplogroup of the Plants' male-line ancestors, underwent a mass migration, around 4500 to 4300 years ago, up the river Danube through central Europe into Western Europe,[52] where the Y-chromosome of the ancestors of the main Plant family underwent a further SNP mutation (P312 also known as S116). The alternative view, based more directly on DNA

[51] John S Plant (2005) *Modern Methods and a Controversial Surname: Plant*, Nomina 28, 113-133, esp. 120, 131-132. John S Plant (2007) *The Tardy Adoption of the Plantagenet Surname*, Nomina 30, 57-84, esp. 80-82. http://cogprints.org/5985/ and http://cogprints.org/5986/

[52] The date of arrival in Europe of the ancestral DNA haplogroup, with the earlier SNP mutation L11 (aka S127), for example, is still disputed and it may have been before or after the Neolithic agricultural transition. Busby et al (2012) Proc R Soc B, **279**, 884-892.

evidence, is that the male-line ancestors of the Plants spread up the Atlantic Coast of Europe to Britain sometime between the British Bronze Age and the late medieval period. A nomenclature for the R-P312+ sub-clade[53] for the Plants is R-DF27+(xZ196,xZ225).[54] A less restricted sub-clade R-P312+(xU152,xL21)[55] is found for example amongst 62% of the population in Spain and Portugal, 29% in France, 13% in Ireland, 11% in Switzerland, but only 6% in England where the main English Plant family is found. This suggests that the Plants' ancestors came from the Continent via France though it remains feasible that they arrived at Macclesfield via Ireland and/or Wales and not necessarily directly from France to England. Further developments of DNA techniques might help in tying down the timescales. Also, finding further SNP mutations in the main Plant family's genetic ancestral line might eventually tie down more specific locations, with relevant timescales, more tightly.

Recent DNA findings suggest more generally a long history of migration up the Atlantic coast of Europe to re-people Britain after the last Ice Age.[56] Aside from the Plants' specific ancestral line, it has even been suggested that there could have been for example an early influx of words from a North Western Semitic language, associable with the Phoenicians, up the Atlantic coast into Britain, that contributed to some early British place names.[57]

Similar notions of Atlantic coast trade might conceivably extend to speculation about the early meaning of *plant*. In Basque, this means `to feign'. This can be considered along with evidence for old Aquitainean deities: these had the names of plants and animals. Conflating the Basque language with early local belief in plant deities, we might note that the feigning of a deity is a concept that is not at odds with the posturings of the medieval nobility. This is consistent with the ninth- and twelfth-century noble names of Plantapilosa (Duke of Aquitaine) and Plantegenest (Count of Anjou).

[53] Earlier testing showed that all of the DNA sub-clades M65, M153, U152, L21 and L176.2 test negative, i.e. these further known mutations have not occurred for the main Plant family.

[54] Here, R-DF27+(xZ196,xZ225) means that the DF27 has tested positive but its sub-clades Z196 and Z225 (and, in fact, some of their sub-clades) have also tested negative.

[55] http://www.goggo.com/terry/HaplogroupI1/y-Haplogroups_I1_and_R1b_in_European_Countries,_plus_Ancient_Migrations.pdf where the sub-clade L21 is called M153.

[56] E.g. Stephen Oppenheimer, *The Origins of the British* (London, 2006).

[57] Richard Coates (2012) A toponomastic contribution to the linguistic prehistory of the British Isles, Nomina, 35, pp. 49-102.

There are also the Middle English phrases of planting vertu, or the Word of God, or noble lineage: these suggest godly or noble aspects to planting relating to shades of difference between creation and procreation. The Middle English word *plente* (sometimes spelled *plante*) meant `fertile' – a concept commonly associated with ancient deities. The fertile enclosure of the Black Prince's vaccary (cattle station), though he was campaigning mostly in Aquitaine (Gascony), can more directly be associated with the contemporary scholastic powers in the medieval plant soul[58] which were nutrition, augmentation and generation; or, more mundanely, feed, growth and breeding. The Black Prince's vaccary is the first known location of the Plants in their main homeland and it is not inconceivable that this fertile enclosure could have been referred to as a "plant" (planting or breeding enclosure) though this predates direct evidence for the Industrial Age usage of this word. As such, the Black Prince's "plant" could conceivably have been the source of the name for a family living there when the Black Prince's administrators at the Macclesfield Court ascribed this pre-existing family a name.[59]

We might accordingly join in the speculation about how the Plant name arose, notwithstanding the comment of Bardsley[60] to wit "I give this up. I can suggest no satisfactory solution". As one specific conjecture, it is not impossible that the main Plant family could have carried on the noble name Planta from the Continent in the twelfth century, perhaps more likely through a cultural rather than genetic influence. This focuses on the possibility, perhaps even a genetic one, of recent origins for the name from afar. On the other hand, to illustrate the possibility of an ancient genetic origin overseas, together with a more recent medieval cultural influence from the Plantagenets, we might consider the specific conjecture that the main Plant family's name could have been ascribed, in fourteenth-century Macclesfield Forest (position 11 in Figure 11), to a family living at or near the Black Prince's vaccary. More generally, the origins of the *de la Planta* and *Plant* surnames might variously have arisen by locative reference to a garden[61] or "planting" enclosure; this leaves no requirement to suppose that those already named Plant (or similar) migrated to the main Plant homeland from elsewhere.

[58] The vegetable soul, according to the thirteenth century scholastics, formed a component of the soul of plants, animals and humans.
[59] This vaccary is conjectured to have been at or near Midgeley and the surname Midgeley was similarly evidently ascribed to another nearby family.

[60] Charles Wareing Endell Bardsley, *A Dictionary of English and Welsh Surnames,* (1872-1896).

[61] Rather oddly, P. H. Reaney and R. M. Wilson(1977), in their *A Dictionary of British Surnames*, give the meaning of the surname Garden as `metonymic for Gardener' even though the early forms *del Gardin* (c1183, Oxfordshire) and *atte Gardyne* (1269, Sussex) directly indicate a locative origin to the name.

## Modelling prolific population growth

To summarise some discussion from the previous sections, it is possible that Plant may be a relatively populous, effectively single-origin surname. That is, all, or almost all, of the people alive today bearing this surname may be connected in lines of familial descent from a single individual who lived within reach of the border between Wales and England in the fourteenth century, or perhaps even further afield even earlier. Moreover, Plant is evidently not unique as a possible large, effectively single-origin surname, albeit that confirmation from DNA studies for other surnames is as yet generally lacking. Irrespective of uncertainties about earlier times, the available DNA evidence suggests the possibility that some fourteenth-century families, including the Plants, had, by 1881, grown to a population in England numbering in the thousands, not including all of the individuals who had emigrated from England and their descendants. One way of testing this proposition is to determine whether such growth is demographically feasible and, if so, under what conditions. In this section we use stochastic modelling in an attempt to address this issue. We shall also, in due course, consider the possibility of a multi-origin component to the name, or an earlier origin.

Much of the variation in population size between surnames can be ascribed to chance variation in the number of surviving sons carrying a surname down to modern times. Indeed, it is a well-known result of the theory of branching processes,[62] which grew out of the study of the population dynamics of patrilineal surnames, that, in a patrilineal society in which no surnames are changed or introduced and the overall population is growing, all of the surnames will ultimately either die out or grow very large. In a population initially containing many surnames, one can expect that after a large number of generations, many of the surnames will have become extinct, many will be very small and on the verge of extinction, and some will be quite populous. Between the fourteenth century and 1881, however, the population of England grew by only a factor of about five. It is thus reasonable to question whether a single family could possibly have grown to a population of over 6,600. We address this question using a stochastic simulation model.

Our model is very similar to that used by Sturges and Hagget[63] and later by King and Jobling.[64] The limitations of these models are well known: they do not allow for emigration or immigration, nor do they allow for changing from one surname to another, nor do we initially include NPEs. Nevertheless, they do provide a reasonable idea of the effect of demographic parameters on population growth. The mathematical details of the simulation are described in Appendix B. It was programmed in the R computer language, and the code is available from the authors. The simulation

---

[62] M A Pinsky and S Karlin, *An Introduction to Stochastic Modeling* (Academic Press, Boston, 2011).

[63] C M Sturges and B C Hagget, *Inheritance of English Surnames* (London, 1987).

[64] T E King and M A Jobling (2009) *Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames*, Molecular Biology and Evolution, 26(5), pp. 1093-1102.

only includes reproductive age males, as will be explained below. It tracks the number of such descendants of a single individual beginning in the year 1311 and ending in 1881.

The starting year was chosen as follows. Like all such models, ours functions in discrete generations. That is, it treats the population process as if a male in 1311 instantly creates a full family. One generation later, every surviving son instantly creates a full family; one generation after that, every surviving grandson instantly creates a full family, and so forth. While this is obviously a gross simplification, it has been found to produce a simulation whose behaviour is surprisingly similar to that of a real population. See Sturges and Haggett for further discussion. Each generation has a length of 30 years (see Appendix B for justification). The year 1311 is the year in the early fourteenth century that differs from 1881 by an even multiple of 30.

Although our analysis is intended to apply to any populous potentially single-origin surname, we will initially focus on Plant. Our interest is therefore in estimating the probability that the combined male and female Plant descendants of a single individual would reach a population of approximately 6,600 in 1881. The most common and most powerful method to address this question is Monte Carlo simulation. This technique is described in Appendix B. In brief, we simulate a very large number of times (specifically, one million times) the line of descent starting from a single individual, and determine the fraction of times that the number of descendants exceeds a specified population value. Although the simulation model includes only males, it does not actually include every male member of the population, but only those who are reproducing in the current generation. In each generation, a substantial number of males will not be actively reproducing because they are too young, and a smaller fraction will be too old. Anthropologists often use the values between one third and one half as an estimate for the reproducing fraction of a human population, and we will use both of these values in order to determine the likely range of their effects on the results. Our model ignores emigration, so that it considers the entire descendant population as having been influenced by growth conditions in England and to be still in England in 1881. The total male population in the census of that year, with the surname Plant, is approximately 3,300. Thus, we may assume that the number of reproducing males with the surname Plant in 1881 was between one third of this (1,100) and one half (1,650). Therefore the objective of our simulation is to determine the conditions in the model under which there are between 1,100 and 1,650 reproductive age male descendants in 1881 of a single active male progenitor in 1311.

As mentioned above, the results of the Monte Carlo simulation provide a means of estimating probabilities associated with the distribution of a surname. For example, the fraction of the one million simulated surname lines whose population exceeds a particular value in 1881 is an estimate of the probability of any real surname line exceeding this value. The fraction of lines that have died out in 1881 is an estimate of the probability that a real surname line will have died out, and the distribution of surname line populations is an estimate of the real distribution in the sizes of surname family populations.

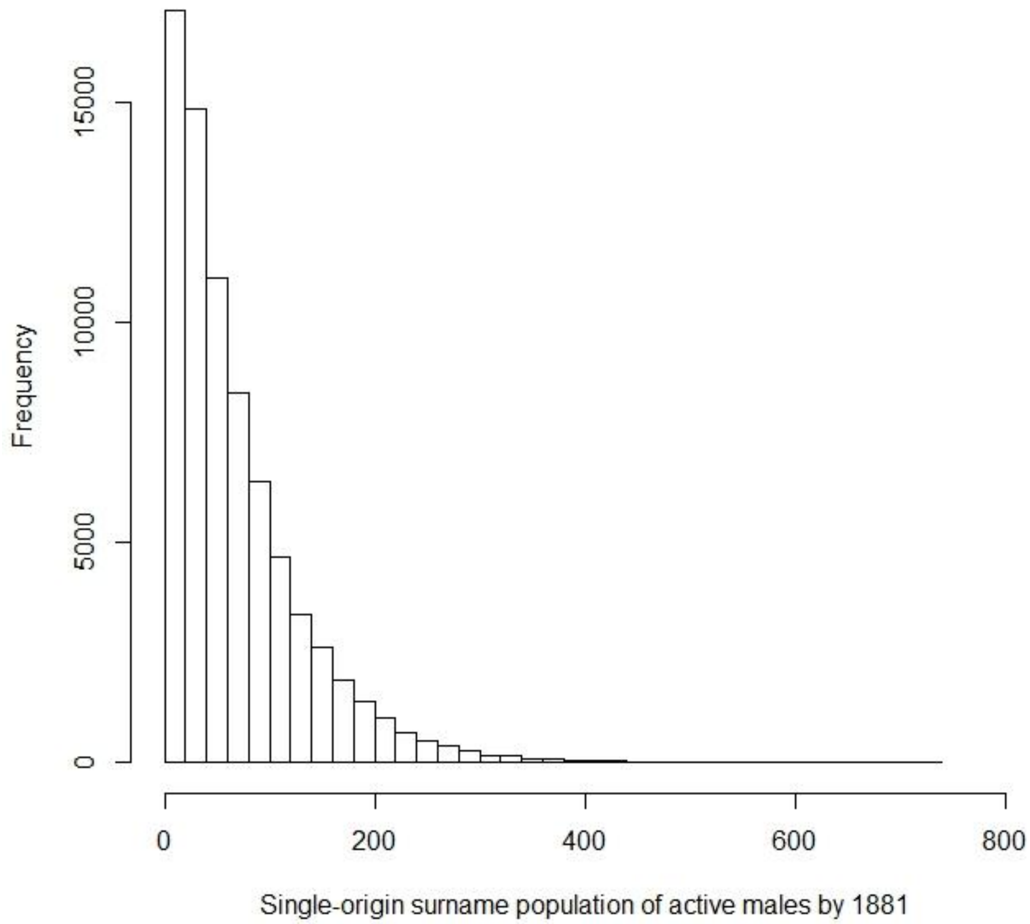



**Figure 15: Histogram of "single family" populations by 1881, under the "base" model using overall population growth data for all of England. Frequency is based on one million initial surname families in 1311.**

## Base model computations

The initial "base" model uses mean population growth rates calibrated from the historical population data of all of England (shown in black in Figure 16). Of the one million "surname lines" initiated in 1311, a total of 75,169, or approximately 7.5 percent, survive to 1881. This is our estimate of the fraction of real surname lines throughout England that survived. Figure 15 shows a histogram of the distribution of the population sizes of the surviving "lines" of descent. Though it cannot be seen in Figure 15, the most populous surname line contains 730 individuals. In other words, none of the one million surname lines achieved a population of 1,100. Therefore we can conclude that it is virtually impossible that paternal lines stemming from a single male in 1311, whose demographics matched the average of the English population, could grow purely by chance to the size of the surname Plant.

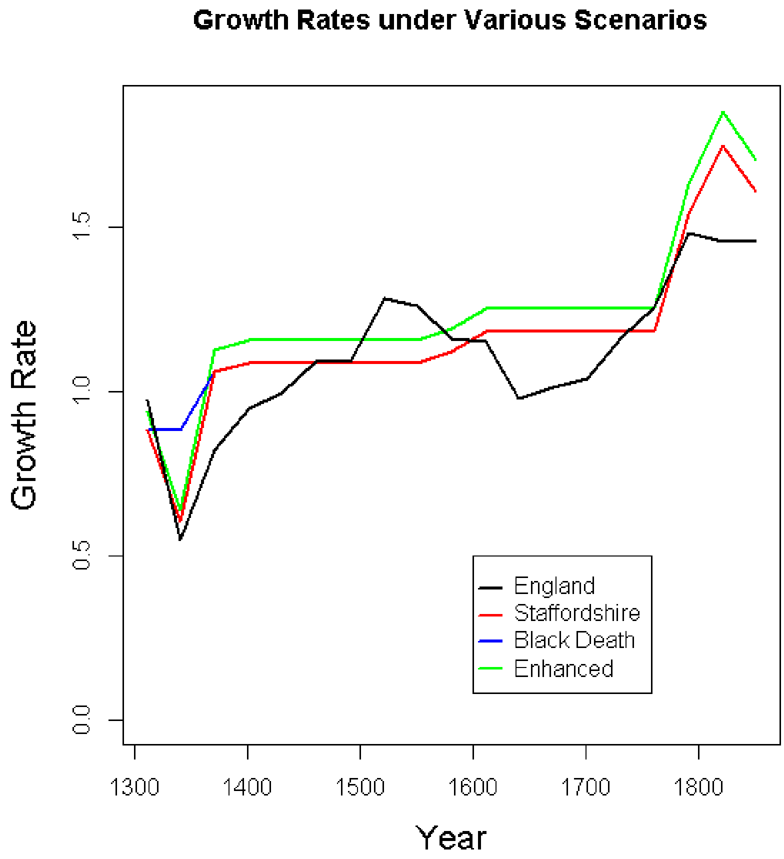



**Figure 16: The base growth rates for the population of England and the higher growth rates for Staffordshire, for an enhanced male reproduction rate, and for enhanced Black Death survivorship.**

As discussed earlier, however, counties in the Northwest enjoyed substantially greater population growth than other parts of England due to the effects of the Industrial Revolution. The Sykes and Plant families, among others, originated in these areas and may have benefited from these effects. In particular, the Plant family arose in southern Cheshire and by the mid fourteenth century had spread into Staffordshire. Our second simulation therefore used population data from Staffordshire. This is shown in red in Figure 16. The results of this simulation were that, of the one million surname lines, three reached a population of 1,100 and none reached 1,650. The population of Cheshire in 1311 was approximately 36,000, which means that there were, roughly speaking, between 6,000 and 9,000 men of reproducing age available to begin a surname line there. Thus we may conclude that it is demographically possible but by no means certain that one of these men could have given rise to a surname line that reached the size of the Plant family in 1881. Two possible scenarios that would increase this likelihood are that the members of the Plant line had some special characteristics, or that Plant is an *effectively* single origin surname with a large main family and some other small ones. We will consider each of these possibilities in turn.

### Four alternative further models

We consider four alternative enhancements to the model using the Staffordshire data.

- The first is an "early polygyny" model, representing the hypothesis that males in a particular family in the population had more than one wife (or mate for whom it was adequately accepted in the community that her offspring could inherit the paternal surname) during one or more early generations.
- The second is an "early start" model, implying that a family had more than one male member in 1311 because this family had begun using a hereditary surname earlier than that year, or because several men in the same paternal line adopted the same name.
- The third is a model, in which a family's growth rate is enhanced during some or all of the generations of the simulation. The specific enhancement we consider is that, for whatever reason, the members of the family were spared the ravages of the Black Death.
- The fourth model involves an enhanced reproduction or survival of male offspring, perhaps due to some sex-linked trait, as discussed by Redmonds et al. in the case of the surname Sykes (Appendix C).

First we consider the simulations of the "early polygyny" model (cf. Appendix D). The model represents the hypothesis that during the early fourteenth century, when the surname was being adopted, one or more generations of males had children with more than one wife, either due to promiscuity or to accepted social custom in that society. The model therefore contains two adjustable parameters: the number of wives per male, and the number of generations in which polygyny occurred. The computations indicate that if the male in the first generation and all of his sons in the second generation have three wives, then the fractions of the one million simulated surname lines that by 1881 reach 1,100, and 1,650 are 0.001, and 0.00003, respectively. Under the "most liberal" assumptions of a reproducing population of 9,000 men and a threshold of 1,100, the number of surnames is nine. Under the "most conservative" assumptions of a population of 6,000 men and a threshold of 1,650, the number is less than one. If the first generation male is the only one practising polygyny, and if he has seven wives, then the corresponding most liberal and most conservative figures are six and zero.

Next we consider the "early start" model. To obtain approximately the same results as those described in the previous paragraph, with approximately 0.1 percent of the simulated surname lines having a population greater than 1,100, the population of the lines in 1311 must be twelve individuals. At population growth rates of the thirteenth century, it would have taken roughly ten generations to accumulate a reproducing male population of this size.

Next we consider a model in which the family's growth rate is not affected by the Black Death. The growth rates for this model are shown as the blue line in Figure 16. The probability of reaching 1,100 was 0.000012, which translates to less than one out of a population of 9,000. Also, under the most conservative assumption, there were no surnames that reached 1,650. Finally, we consider the enhanced male hypothesis. The simulation results indicate that if the fraction of males born into the family is enhanced by a factor of six percent (cf. Appendix C and green line in Figure 16), then the probabilities of reaching 1,100 and 1,650 in 1881 are 0.001 and 0.00017, respectively.

Figure 17 shows the growth of means of the largest one thousand (i.e., the largest 0.1 percent) of the simulated lines under each of the four enhancement scenarios, together with the corresponding mean of the unehanced simulation. The enhanced Black Death survival scenario does not provide sufficient growth to reach a population of 1,100 and can likely be rejected as a sole cause of the large 1881 population size. Among the other three scenarios, the Y chromosome linked male enhancement tends to achieve a larger size than either the early polygyny or early start scenario. This is due to the enhanced reproductive rate during the entire period (Figure 16).

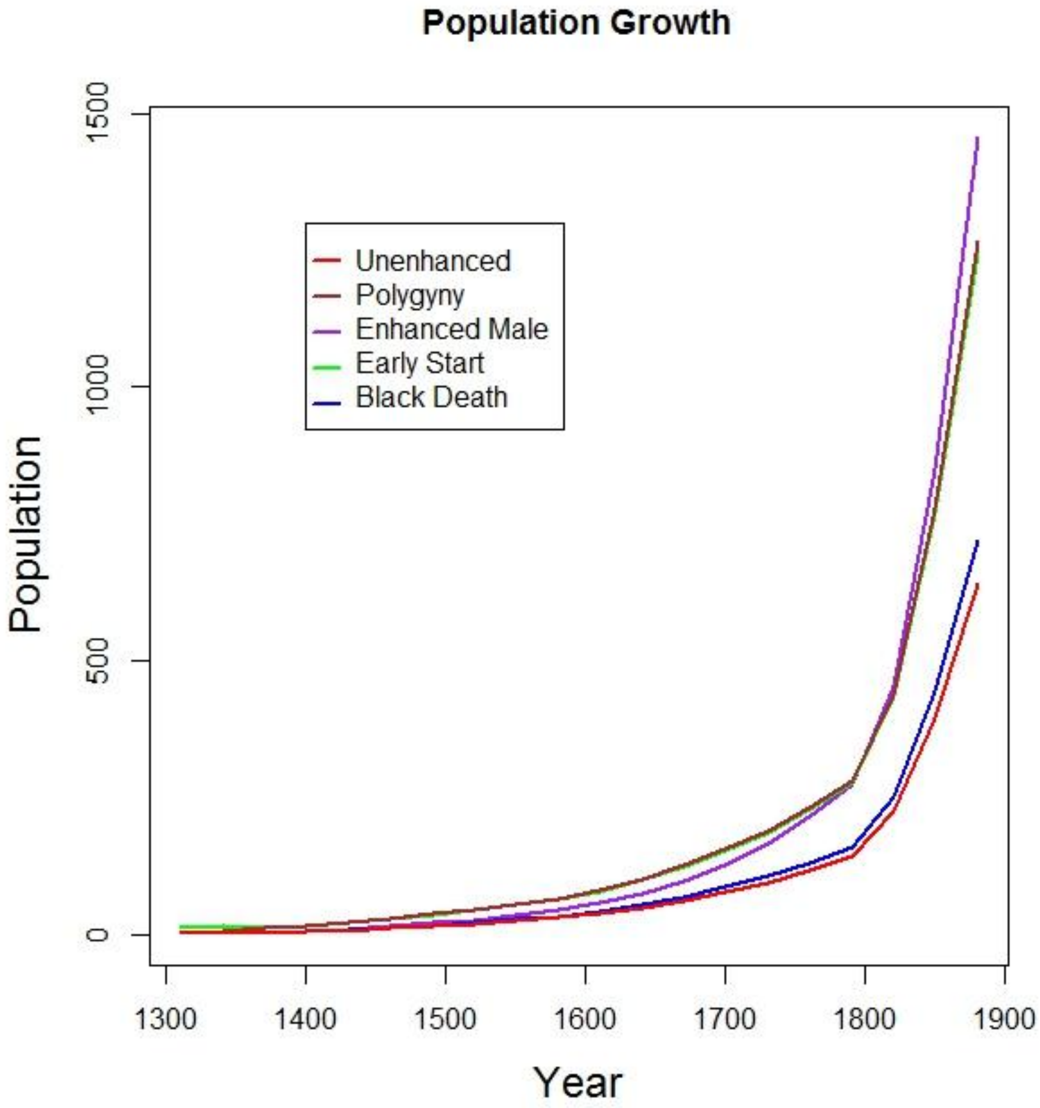



**Figure 17. Population growth curves of the means of the largest 0.1% of the simulated lines in each of the four enhancement scenarios, together with the corresponding mean using the unenhanced Staffordshire parameters.**

## Effects of county growth rates

The discussion surrounding the empirical data in the chart in Figure 6 indicates that there are substantial differences in population among the most likely single-origin surnames in the six counties whose data are shown in that chart. Part of this difference can be attributed to different population growth rates in these counties due to regional differences in the effects of phenomena such as the Industrial Revolution. It is instructive to compare the results of simulations employing the growth rates of these counties. These growth rates are shown in Figure 18.

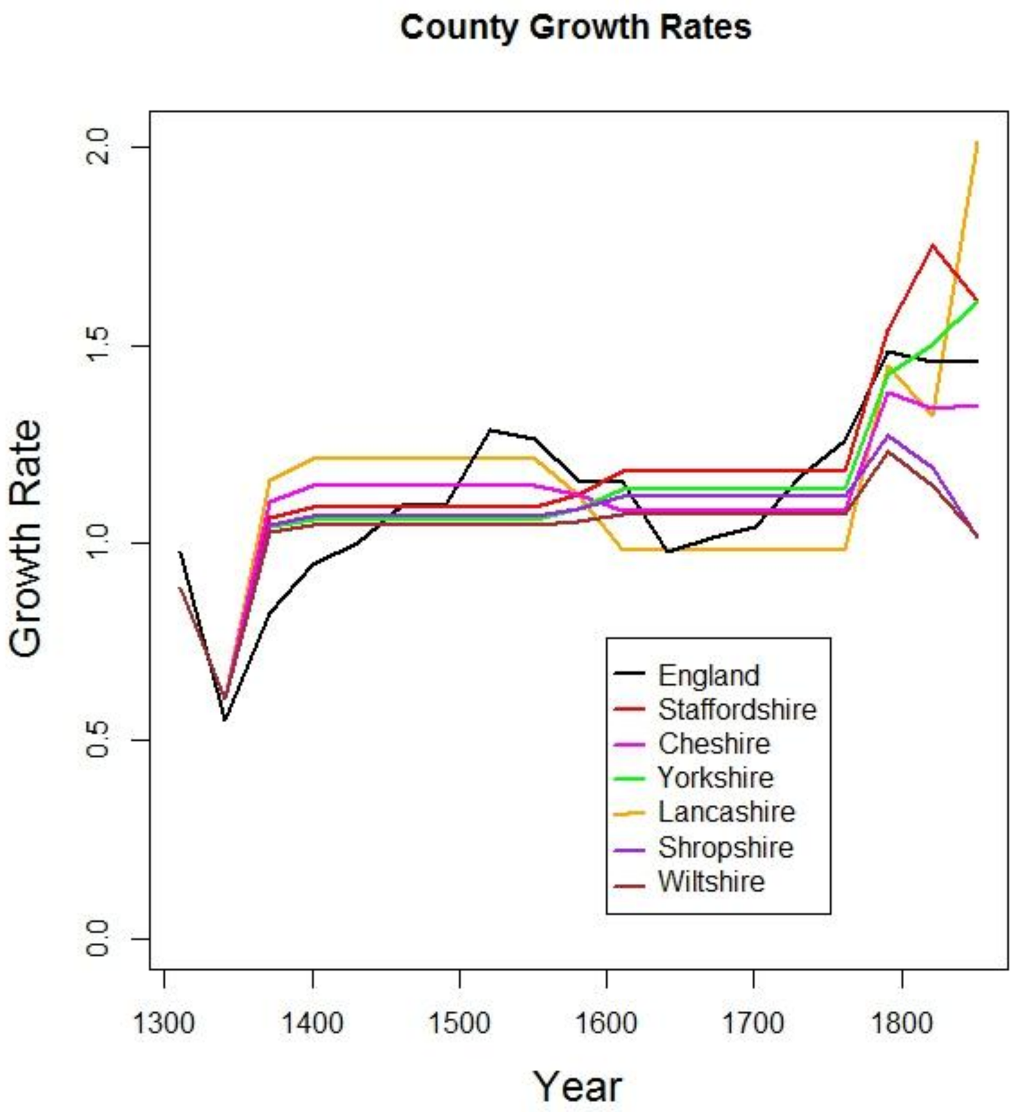



**Figure 18: Population growth rates for the six counties in Figure 6.**

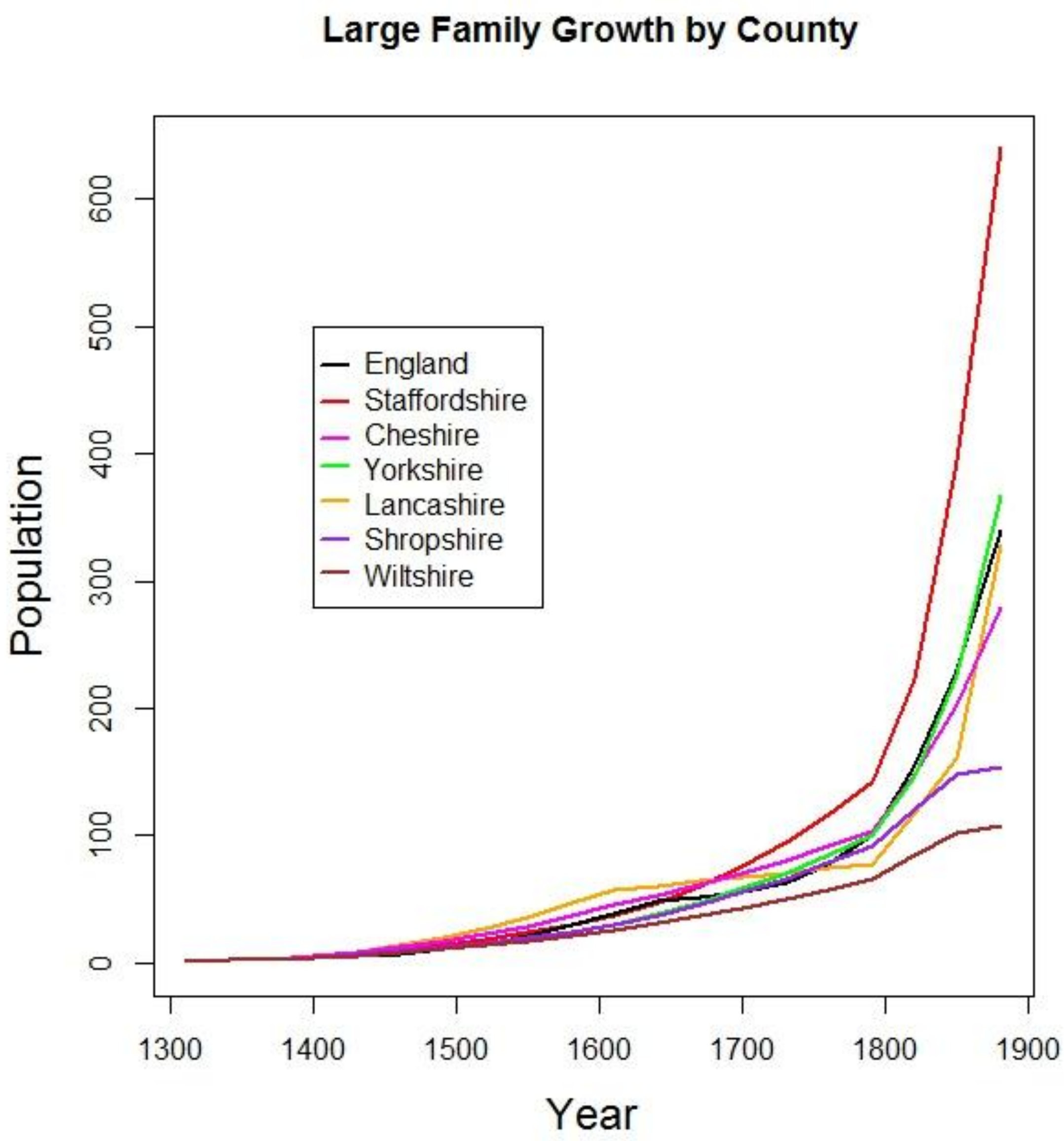



**Figure 19. Population growth curves of the means of the largest 0.1% of the simulated lines in each of the six counties, together with the corresponding mean using the base England parameters.**

Figure 19 shows the population growth curves of the means of the largest 0.1% of the simulated lines in each of the counties. Table 1 shows the largest simulated populations in 1881, in simulations using these parameters. It must be emphasized that these are the largest populations in Monte Carlo simulations of one million lines, and thus do not represent simulations of the actual counties but rather estimates of the maximum population that could conceivably be obtained purely by chance within the county demographic parameters.

| Yorkshire | Lancashire | Cheshire | Staffordshire | Shropshire | Wiltshire |
| --- | --- | --- | --- | --- | --- |
| 831 | 658 | 577 | 1246 | 332 | 229 |

**Table 1. Largest simulated populations by county.**

The results in Table 1 show, in particular, larger family sizes by 1881 in Staffordshire than for Cheshire, Shropshire or Wiltshire and this agrees well with Figure 6. However, the Table 1 maximum family sizes differ from those of Figure 6 in that the largest simulated populations in Yorkshire and Lancashire are considerably smaller than one sixth to one quarter of the most populous single-origin contenders shown in Figure 6. This could be because these Lancashire and Yorkshire surnames are not actually single-origin, due to some local quirk of naming convention in this region of West Yorkshire and south east Lancashire in particular, but it could also indicate that rapid population growth of some families might have been due to local conditions that are not reflected in the overall county population dynamics. Simulating at a smaller spatial scale would probably be fruitless because it would be impossible to separate out the effects of migration from local family growth.

## Summary of the simulated effects

The models discussed so far indicate that it is possible but by no means certain that a single progenitor in Cheshire in 1311, whose family moved in the fourteenth century to Staffordshire, could have given rise to the entire Plant family, but that this is most likely to have occurred if there were special circumstances, such as early polygyny, an early start, or enhanced male reproduction, that characterised this family. The alternative, discussed earlier, is that Plant is an *effectively* single source family; that is, that one branch of the family is so dominant that it is difficult or impossible to identify other branches. Because the vast majority of surviving surname lines in any population have quite small populations, these secondary family branches would most likely be much smaller than the "main" branch and might be very hard to identify or to distinguish from the effects of NPEs.

To estimate the relative contributions of small surname branches, from separate origins, we carried out a Monte Carlo simulation in which we randomly selected nine simulated surname lines from the 103,117 surviving lines of the unenhanced Staffordshire simulation and aggregated them to the largest line. On average, the line of the largest simulated family made up 54% of the aggregated population for the surname. This contribution of the large family diminishes if there are a *very* large number of smaller lines in the surname: when this test was repeated using ninety-nine randomly selected lines of separate families, the main family line made up 9% of the aggregated population. These estimated contributions are probably low, since the small families could in principle come from anywhere in England rather than only from the favourable conditions of Staffordshire. When the simulation was repeated using values from all of England for the small families, the portion of the simulated population from the main family line was 67% if there were nine small families and 15% if there were ninety-nine small families.

## Discussion of the evidence for Plant and Sykes

It is possible that the population of a surname, originating with one male in the early fourteenth century, could expand to the size of Plant or Sykes as observed in 1881, but to do so would require a combination of genetic drift (i.e. the chance production of more male offspring) and some sort of favourable circumstances not enjoyed by most of the population of a county.

If a very large number of separate families in a surname survive into the modern era then it is almost impossible that one of these will dominate the others to the extent observed for Plant and Sykes. The situation in this case will be more like that of Smith, which is usually described simply as a multi-origin surname. It hence seems appropriate to simulate the unusual situations for Plant and Sykes by means of an unusually dominant family and a relatively small number (perhaps of the order of five to ten) of more typical surviving families, which are much smaller than the exceptional dominant family. The histogram of Figure 15, which is an estimate of the distribution of surname populations in 1881, indicates that, just due to chance, a few very large surname families will develop. The specific identity of the surnames that will be fortuitous in this way is purely random; but, in the event, it appears that two of these surnames could have been Plant and Sykes.

As a specific example, we will discuss the case of Plant, before moving on to Sykes.

The medieval data in Figure 11 indicate twelve distinct possible origins for the Plant name and there are others[65] not included in the Figure, not to mention probably still more that have not been found. The simulations suggest that only around 10% of these can be expected to have survived but this still implies that we can consider that there could well be a few surviving small Plant families with distinct origins in England as well as the abnormally large Plant family based in Cheshire and Staffordshire. These more typical small families may have originated with different meanings in different places before simplifying to the spelling of the common English word *plant.* For example, in Sussex, there are the names *Plant* and *Pallant*, which may have led to some confusion of pronunciation and spelling, and there are also several early instances of the name *Plente*; all of these have different meanings and this is before we consider the many different meanings of the Middle English, French and Welsh word *plant* not to mention such early name forms as *de la Planta* and *de Plantes*. We are sceptical that the meaning of the main Plant family's surname is simply `gardener', as is often stated. This is not least because the name *Gardiner* appears along with *Plont* in the main Plant homeland, in the Black Prince's fourteenth-century Court Rolls for Macclesfield in Cheshire, so that we may raise the question of why should we suppose that the word *plant* was used indirectly for a gardener, in the main homeland, when there are other possible appropriate meanings to this

[65] http://www.plant-fhg.org.uk/origins.html#13c

word such as children or (from the) planted enclosure. In short, the semantics of the Plant surname is of little or no help in ascertaining the origins of this name.

Our contention of more than one surviving family called Plant is reinforced by the 1601-1850 baptism data for Plant. It can be seen in Figure 13 that the Plant baptisms along the east coast of England, in Lincolnshire, Norfolk, Suffolk and Essex, form a geographically rather separate block from those around the main Plant homeland of Cheshire and Staffordshire. We might add that, consistently throughout these dates, the baptisms in Cheshire and Staffordshire account for only about a half of the total number of Plant baptisms, though the DNA matching data in Figure 14 show that some of the matches to the main Plant family have spread somewhat to the east of Cheshire and Staffordshire. However, such easterly matches are not as yet found as far east as the east coast of England; in Lincolnshire for example, there are a pair of DNA matches to a different "south Lincolnshire Plant family" which seems likely to have descended from a different origin.

Considering the set of independent Plant DNA volunteers who have had at least 12 markers measured, 23 out of 34 have matched. If we restrict our consideration to those whose ancestral male-lines are known to have been in England in 1881, this reduces the result to 9 matches out of 17. This latter `England-based' result is an attempt to remove the possible effects of different conditions overseas, since a few emigrants to America (or indeed to Ireland or Australia) might have gone on to experience very different growth conditions to those that we have simulated in England. Thus, for example, a few matching emigrants to America might have grown to large descendant families with disproportionally many matches amongst those tested. The match rate is 82% (14 out of 17) for those Plants whose male-line is not known to have been in England in 1881 as against 52% (9 out of 17) for those who were. As well as reducing the fraction matching from 62% to 52% for our Plant study, this `England-based' result significantly reduces the size of the DNA dataset from 34 to 17 and accordingly increases the standard statistical error for the matching percentage from 8% to 12%. For a populous surname, such as this, the standard error when extrapolating to the whole surname depends dominantly on the *number* tested, unlike the situation for rare surnames where it depends on the *fraction* tested.[66] Besides considering the standard error, we have here adopted an approach of seeking to eliminate both non-randomness and sources of possible bias in our dataset. Though our DNA dataset is small, especially when emigrants are omitted, we may surmise that the DNA results remain consistent with the more general contention of a dominant Plant family around Staffordshire.

[66] Though James M Irvine (2010) *Towards improvements in y-DNA Surname Project Administration*, Journal of Genetic Genealogy 6(1), 24 pages http://www.jogg.info/62/files/Irvine.pdf stresses the use of a penetration factot, which he defines simply as the fraction of the surname tested, he does not make this important distinction between frequent and rare surnames.

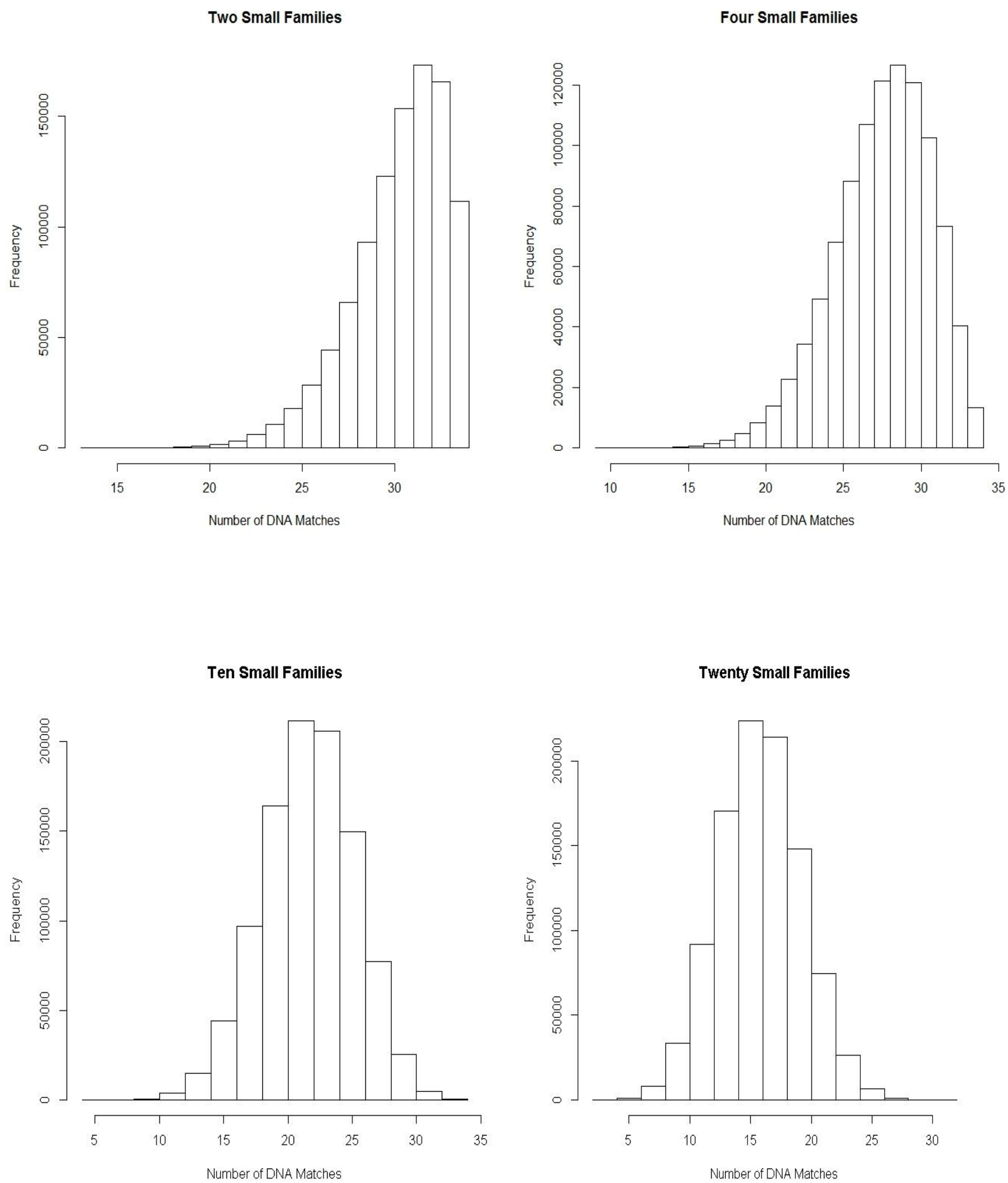



**Figure 20: Model calculations of the chances of different numbers of DNA matches for the 34 tested Plant men in the full independent-volunteer dataset.**

Considering first the larger Plant dataset, the model calculations indicate that the best chance of there being an experimental result of 23 DNA matches, out of 34 Plant volunteers, arises if, neglecting NPEs, there are around ten typically-small families in addition to the dominant Plant family. This is represented by the bottom left histogram in Figure 20 for which the peak frequency corresponds to around 22 DNA matches. The chances of observing the experimental result of 23 matches for Plant are smaller for other numbers of small families (i.e. there are smaller frequencies corresponding to the value 23 on the x axis in the other histograms of Figure 20). Here, the typical families, which are relatively small, are drawn at random from the computer simulation results for

Plant that are based on the population growth factors for England whereas the dominant family is based on exceptional fortuity within the growth conditions of Staffordshire. However, this neglects the effects of false paternity events (NPEs) and, from the documentary evidence for Plant, we might instead expect that there are around four typically-small surviving Plant families, in addition to the dominant one, with the 11 mismatching DNA results being explained partly by these separate origins and partly by NPEs. If we now consider instead the smaller `England based' dataset, it can be seen that there is rather more scope for different numbers of small families leading to the experimental result of 9 out of 17 DNA matches though, in the absence of any NPEs, around 10 to 20 small families would suit (Figure 21). We will consider the effects of NPEs more explicitly in a further paper that we are preparing but the important point to make here is that the 11 DNA mismatches, for the 34 tested men, or 8 in the bias-reduced dataset, may not be entirely due to NPEs and that the DNA results do not hence prove that Plant is a single- rather than a plural-origin surname.

If we were to presume that there were for example a dominant single Plant family and as many as ten typically-small surviving Plant families that have different origins, with relatively few NPEs, we can expect that about two thirds of the Plants in England will have descended from the main Plant family with the remainder coming from the smaller families. In other words, our observed DNA result for Plant is a possible one if there are about ten small surviving Plant families in addition to the dominant one with few NPEs. It is important to add that the Plant DNA results can instead be explained as just one abnormally large single family with the 11 mismatches (or 8) due solely to NPE-descended men amongst the 34 (or 17) independent DNA volunteers. However, if there were ten surviving typical Plant families in addition to the dominant one, the smaller separate-origin families would account for perhaps one third of the total Plant population and this would reduce the target population for the main Plant family from the range 1100-1650 active males to the rather less demanding range of 770-1100. This consideration that the name may be significantly plural-origin is not essential for Plant, since its dominant family can be explained as one of the most fortuitously growing families in the general growth conditions of Staffordshire, with only a small need to assume further advantageous factors or to assume a significant population from its plural origins; most of its exceptional size can be explained if the main Plant family simply had extremely many fortuitously surviving sons in the favourable growth conditions of Staffordshire. On the other hand, an explanation for the larger population of the Sykes surname is more elusive, if it is taken to be single-origin as has been claimed, and it is accordingly more imperative to consider that it could be plural-origin.

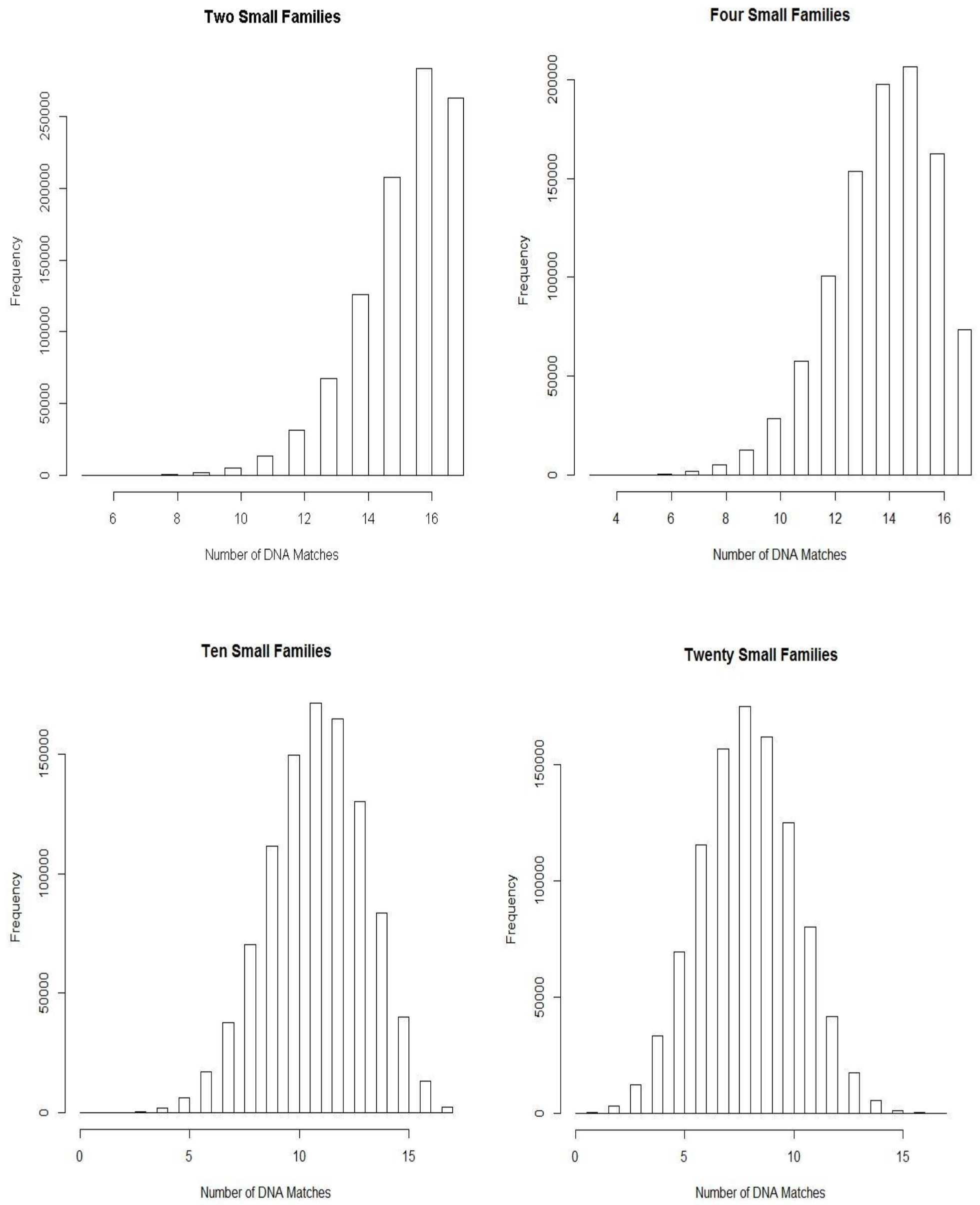



**Figure 21: Model calculations of the chances of different numbers of DNA matches for the 17 Plant men in the `bias-reduced' dataset.**

If Plant were purely single-origin, the NPE rate derived from the DNA result would be 1.6% per generation, or it would be 2.6% in the case of the reduced `England-based' dataset. Sykes has a rather higher fraction of DNA mismatches leaving more scope for it being plural- rather than single-origin. The usually quoted NPE rate for Sykes is 1.3% per generation, which is copied from the Sykes and Irven paper,[67] but this is apparently based perhaps on an average value from some forward simulations and, when the more usual back-method of calculation is used, the Sykes NPE rate is

[67] B Sykes and C Irven (2000) *Surnames and the Y Chromosome*, Am. J. Hum. Gen. 66(4), pp. 1417-1419.

3.53%. In other words, if Sykes is assumed to be a single-origin surname, the deduced NPE rate is quite high and it seems likely that Sykes has in fact plural origins and a lower back-calculated NPE rate than this 3.53% value. The presence of a few small separate-origin families would reduce the implied size of its dominant single family.

The size of the dominant Sykes family is deduced from the size of the total Sykes population. In the UK, the total population of Sykes in 1881 is 14,383 but the Sykes and Irven experiment considered only West Yorkshire, Lancashire and Cheshire, which reduces the total Sykes population here to around 11,000; of these it is estimated, with the usual factors of a sixth or a quarter, that around 1,800 to 2,750 were active males. As a proxy to growth conditions for Sykes, we have used the historical growth factors for Lancashire and found that the most fortuitous family, along with a few small surviving typical families, can barely reach these targets of active Sykes males. As many as 20 small surviving families are needed and then only a small fraction 0.024 of the simulations reaches the lower target of 1,800 active Sykes males and, beyond that, there is only a tiny chance 0.000004 of the simulated number of active males reaching 2,750. It would appear almost certain that additional beneficial factors are needed to explain the apparently large population of the dominant Sykes family in this area in 1881, beyond its exceptional fortuity in bearing sons, due to random chance, even with 20 assumed separate-origin surviving Sykes families and the general growth conditions in Lancashire.

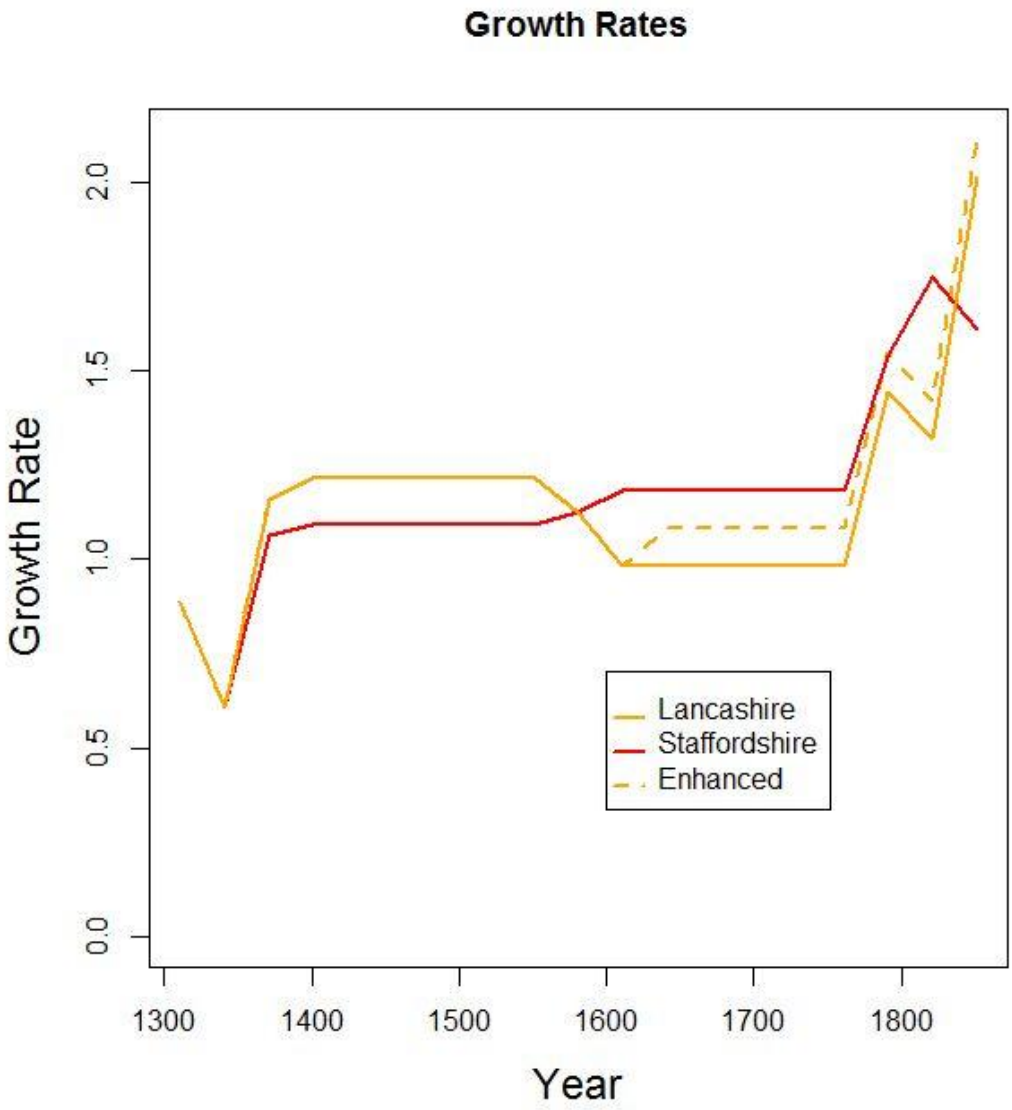



**Figure 22: Population growth rates relevant to Plant and Sykes.**

It appears that the dominant Sykes family originated in West Yorkshire where wool wealth was prevalent from early times leading on into the Industrial Age. Precise figures for the historical growth rates of the population under these favourable conditions are not available but we might, in the absence of better data, assume that they were similar to those of Lancashire with the addition of 0.1 to the values between 1671 and 1851 (Figure 22). For *ten* small families we then get the probability 0.041 of exceeding the 1,800 target population and 0.0002 for exceeding 2,750. Thus, by adding this moderate value to the population growth factors, a feasible model is obtained for explaining the Sykes DNA result, with ten small families which leaves scope for NPEs (Figure 23).

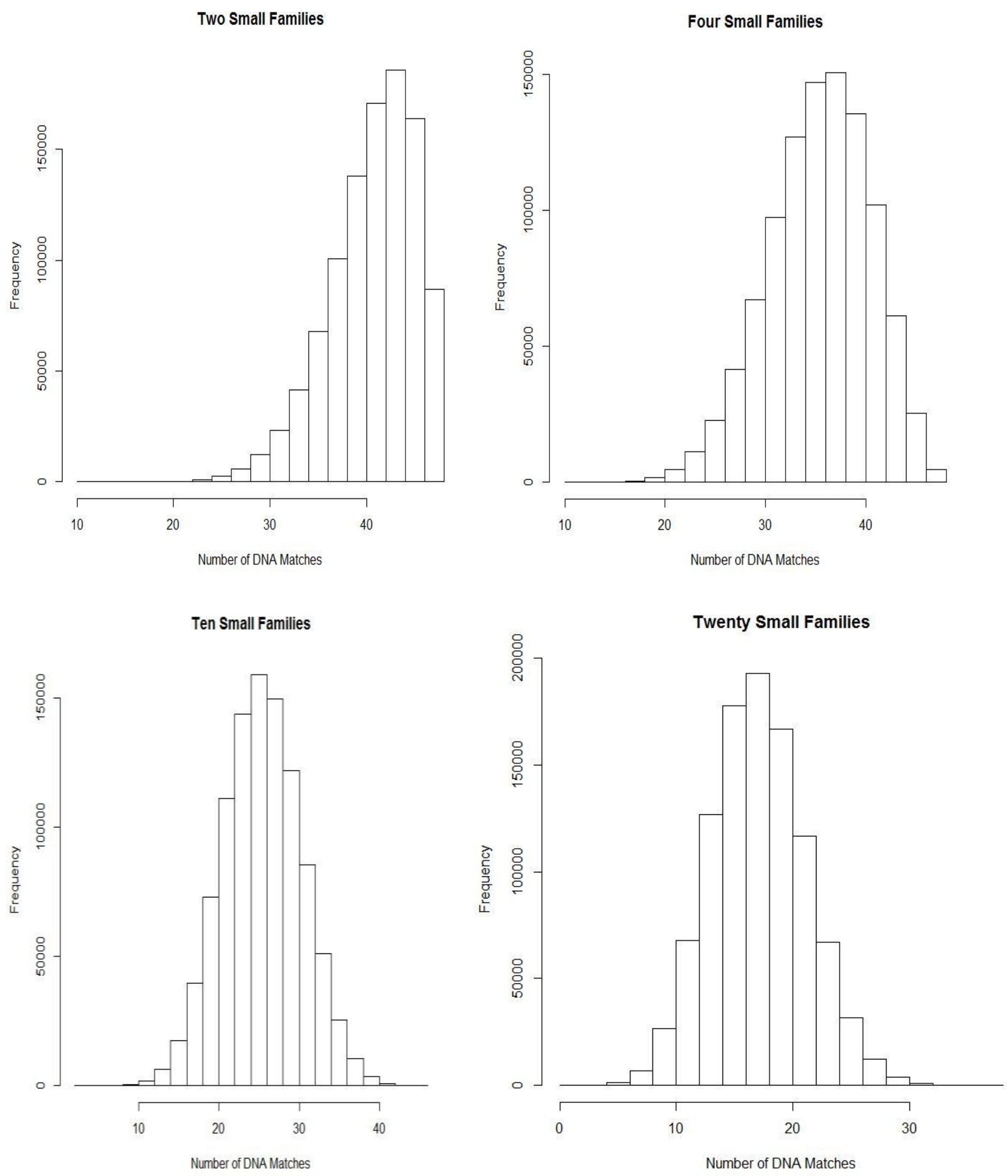



**Figure 23: Model calculations of the chances of different numbers of matches for the 48 Sykes men tested in the initial DNA study.**

It is often claimed that it is fallacious that there is a large apparent size to the dominant Sykes family since this published assertion is based on data that were unsound because the Sykes and Irven DNA study was very low resolution (only 4 Y-STR markers were measured). It is claimed that the conclusions of the published Sykes and Irven study were misleading for this reason. Overstatement of either the strengths or the weaknesses of the study should be avoided however. The Sykes and Irven paper states, “This [Sykes] haplotype is not encountered in either the Neighbor or English controls”. This implies that it is unlikely that the 21 matches for the 48 tested Sykes men were false positives; they cannot be dismissed simply by pointing out that the study was low resolution. That is not to say that other researchers have not strived to do better; the Sykes and Irven study made no attempt to remove people from their dataset who were already known to be closely related, as

K&J[68] for example did for other surnames; also, there is invariably the drive for larger sample sizes given, for example, that their non-Sykes "neighbour" sample contained only 21 males. It may be questioned, not only in particular whether it was that all of the 27 observed DNA mismatches were due to NPEs, but also whether the published results were representative of the whole Sykes population, especially as it sampled only three counties. We accordingly have considered *only* that there appears to be a dominant Sykes family within the Sykes population of West Yorkshire, Lancashire and Cheshire which is consistent with the 1881 surname distribution data (Figure 6).

Our model simulations for the expected number of DNA matches amongst the 48 tested Sykes men is shown in Figure 23. These simulations assume that the 27 mismatches are all due to separate origins to the surname and it is important to stress that some are no doubt due to NPEs instead, not to mention possible bias in the sample or other errors. However, taking the results at face value, there remains scope for NPEs as well as ten small families and still obtaining an experimental result of 21 out of 48 DNA matches, still bearing in mind that there is a statistical standard error of plus or minus 3 on the 21 matches not to mention possible systematic errors in the much contested published result. NPEs can be expected, in effect, to displace the model results for ten small families (lower left histogram in Figure 23) further to the left still coinciding adequately with the published experimental result of 21 matches, with further leeway due to the experimental uncertainties in the result.

## Conclusions

Newly available DNA results at the turn of the millennium purported to confirm George Redmond's opinion that some West Yorkshire surnames,[69] such as Sykes, were single-origin despite being populous. Richard McKinley made similar intimations[70] concerning apparently large single families in neighbouring south-east Lancashire. One of the authors, Bryan Sykes, of the initial Sykes DNA study[71] suggested the explanation (Appendix C) that the apparently large single-family of related Sykes men might all possess an inherited Y-chromosome that leads to an enhanced tendency to sire more sons. However, there are other possible explanations.

[68] T E King and M A Jobling (2009) *Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames*, Molecular Biology and Evolution, 26(5), pp. 1093-1102.

[69] George Redmonds, English Surnames: Yorkshire West Riding v. 1 (English surnames series, 1973)

[70] Richard McKinley, Surnames of Lancashire (English surnames series, 1981)

[71] B Sykes and C Irven (2000) *Surnames and the Y Chromosome*, Am. J. Hum. Gen. 66(4), pp. 1417-1419.

The initial DNA study for tested Sykes men was for volunteers from West Yorkshire, Lancashire and Cheshire. The estimated total 1881 population of people named Sykes in these three counties was around 11,000. Of these, around 1,800 to 2,750 can be expected to be males actively engaged in producing offspring. The Y-DNA results for Sykes displayed 43.8% matches, signifying a dominant single family. The remaining mismatches were attributed entirely to NPEs which arise from occasional females passing down the surname, breaking otherwise intact male lines of descent of both the Y-chromosome and the surname. However, since typically-sized families are small (Figure 15) in comparison to a dominant one, they can be expected to lead to little structure in the array of mismatching DNA Y-signatures, making them essentially indistinguishable from the usual presumption that quite substantial levels of mismatches can be attributed to NPEs. If there were, instead of many NPEs, relatively few but instead many other origins, as might be the case, the Sykes DNA results would allow as few as around 900 to 1,400 active males in the 1881 population of the *largest* Sykes family, with the remainder attributable to more typically-sized small families from other origins.

Taking typical population growth parameters for England as a whole, our computer simulations indicate that the largest family possible, due to pure chance alone, is around 730 active males (Figure 15). If we take population growth values for Lancashire, as perhaps more typical of the conditions pertaining for the dominant Sykes family, then the predicted largest family has a slightly smaller size of only 658 active males (Table 1). This value is increased to 1,270 active males, however, if the Lancashire growth factors are enhanced during the times when some families in West Yorkshire and south-east Lancashire were benefiting from wool wealth leading on into, and through, an early birth to the Industrial Age. In order to obtain the required result for Sykes, we have enhanced the county-wide Lancashire growth rate values by 0.1, between 1671 and 1851 (Figure 22). This then brings the largest predicted family in 1881 to 1,270 active males, which is within range of the 900 to 1,400 needed, if Sykes is plural origin with many typically-small families besides its dominant one.

Explanations based on an idea of favourable population growth conditions in particular regions hold some general attraction for populous single-origin surname contenders. The 1881 geographical distributions of surnames (Figure 6) suggest that there might be several particularly populous families in West Yorkshire and Lancashire, perhaps similar to Sykes, though this has not been confirmed by DNA evidence. It seems that Sykes is the second most populous West Yorkshire surname whose geographical distribution in 1881 seems consistent with the dispersal of a dominant family from a single source (Figure 4). In our theoretical computations, different county-wide growth parameters can lead to diverse predicted sizes for the largest single-family populations by 1881, varying from 1,246 active males for Staffordshire to 229 in Wiltshire (Table 1). Up to a point, theory (Table 1) and observation (Figure 6) are in accord in that, for both, there are very diverse sizes of the largest single-origin contenders in these six counties. The underlying county-wide growth factors (Figure 18) for the theoretical computations are based on published documentary data (Appendix B). However, in order to explain the particularly large sizes of the contenders in West Yorkshire and

Lancashire, it has been noted that it has been necessary to enhance the published county-wide growth factors. Even growth parameters for regions as large as a whole county might be subject to the effects of migration such that they do not entirely reflect the growth of the native population; they are accordingly subject to some uncertainty for our purposes. There is hence only indirect evidence, which is feasible but not yet numerically demonstrable, to support the theoretical 0.1 growth-factor enhancement that we have adopted to explain the Sykes Y-DNA results.

The maximum values in Table 1 represent a one in a million chance for a typical family of active males in each county growing this large. In fact, it is perhaps more realistic to consider the mean population of the largest several simulations, computed for each county, as is shown in Figure 19. This substantially reduces the simulated large-family size, from 1,246 to around 670 active males for Staffordshire for example. However, this assumes that all families in the county are subject to identical conditions whereas some may have benefited from more favourable conditions than others, perhaps in particular regions in a manner related to the local economy for example. For example, it is evidenced at Halesowen in Worcestershire, around 1300, that the wealthier peasants had an average of 5.1 children as against 1.8 for the cottagers; also, mortality during the Black Death could be as low as 27% in the case of tenants in chief of the English crown but as high as 70% for some peasant groups; similar considerations probably applied to years of famine.[72] In matters such as wealth, there can be expected to be a variation amongst the individuals within a populous family but some particular beneficial factor might pertain in a family's particular homeland, perhaps persisting on into more modern times. There is data to show that, between 1761 and 1841, population growth in some semi-rural Wapentakes of West Yorkshire and south east Lancashire and east Cheshire was exceptionally high in comparison to the overall trends in their respective counties;[73] this might perhaps be related to yeoman farmers in these areas benefiting from extra income upon taking advantage of early industrialisation. These latter areas seem partly coincident with the exceptionally large families, such as for Sykes, that have been contended to exist in parts of West Yorkshire and Lancashire in particular (e.g. Figure 6).

More directly from published statistics, the overall population growth factors were particularly favourable in Staffordshire (Figures 18 and 19) though it cannot be ruled out that some of this might be due to migration into the county from elsewhere, such as from Cheshire into the Potteries in north Staffordshire in relatively modern times (Figure 18).[74] Our computations show that the published population growth for Staffordshire would allow an exceptionally fortuitous family there

[72] Christopher Dyer, *Making a living in the Middle Ages: the people of Britain 850-1520* (New Haven and London, 2002), pp. 158, 161, 271-2.

[73] http://www.hpss.geog.cam.ac.uk/research/projects/occupations/englandwales1379-1911/figure2/figure2b.html

[74] Further to the south Birmingham is in the historic county of Warwickshire with some suburbs in Staffordshire and Worcestershire.

to grow, just by random chance, potentially to a size as large as 1,246 active males by 1881 (Table 1), as a one in a million event under average county-wide conditions and , to this result, we can add that Figure 8 suggests Plant is empirically the second most populous single-origin contender in Staffordshire. This surname requires around 1,100 to 1,650 active males to explain its total 1881 population; this observed range is broadly in accord with the theoretical computation of 1,246 even though this agreement allows that the Plant surname is wholly single-origin. Besides other uncertainties, the Plant surname might be slightly plural origin, which could help bring theory and observation further into line, though only a small number of typically small Plant families, in addition to the dominant one, appear to be necessary to explain its observed DNA results (Figure 20). If we consider a reduced `England based' dataset, the fewer DNA results tell us rather less about the number of typically small families needed, assuming few NPEs, though the results remain consistent with around 5 to 10 surviving families which has been inferred from the medieval documentary evidence. The size of the dominant Plant family can be explained largely just by exceptional fortuity within favourable county-wide growth conditions in Staffordshire, with relatively little contribution from other origins, and this leads to a comparable single-family size for Plant to that which we have suggested *might be* the case for largest Sykes family, given that we can allow that the Sykes surname might perhaps incorporate a more substantial number of smaller Sykes families. However, this does not rule out that the dominant Sykes family is in fact larger than this would imply, though that would then require some further beneficial factors for the theoretical growth of the dominant Sykes family. We have already mentioned the feasibility of exceptionally high regional growth in the main Sykes homeland. We have also considered various other factors, more explicitly in connection with Plant, for which theory and experiment are brought more readily into accord; for Sykes, we are so far hampered by a lack of good population growth statistics for the specific region of West Yorkshire and south-east Lancashire, though the aforementioned exceptional population growth here, known in particular Wapentakes for 1761 to 1841, helps in providing a possible qualitative explanation, albeit not yet on the basis of more complete numerical data.

The all-Staffordshire computer simulations leave scope for additional factors that might bring them more fully into line with the apparent single-family population for the Plant surname. These include early polygyny, an early start to the surname's single-family stock, or genetically-enhanced male-line fecundity. Resilience to the Black Death (Figures 16) does not have much of an effect on a single family's growth (Figure 17) in comparison to these other factors. A more substantial effect arises (Figure 17) if the progenitor of the surname family had seven wives or recognised mistresses, or if he and his sons each had three such concubines. A similar effect arises if the stock of males taking up the surname derives from a common ancestor dating back to around the year 1011 such that there were around twelve bearers of the surname by 1311. Surnames are not generally thought to have stabilised before the fourteenth century but we might perhaps envision, for example, a related stock of men living near a geographical feature, such as some ditches (Sykes) or a planted enclosure (Plant), who all acquired the same surname. Though it is generally presumed that the authorities gave the peasants distinct surnames, there might be exceptions to this. A slightly greater effect arises with a Y-chromosome linked fecundity, if that leads to six percent of extra males above the norm surviving to produce offspring at every generation; such a scenario, though not confirmed, has not been ruled out by baptism and births data (Appendix C).

To summarise, for the growth of populous single families, exceptional fortuity in bearing many sons purely by chance is a primary consideration. On this basis alone, a family's size can vary from zero to many hundreds of active males. Beyond that, different population growth factors in different regions appear to play an important role such that, with simple assumptions, by 1881 a maximum family size of 229 active males can be computed for Wiltshire as against 1,246 for Staffordshire (Table 1). These values can be multiplied by between four and six to give the total population of a single-origin surname in 1881 and this is broadly in agreement with the observed evidence (Figure 6) except for Yorkshire and Lancashire, for which our computer models have not explicitly included exceptional population growth conditions in particular Wapentakes and Hundreds of West Yorkshire and south-east Lancashire. It is furthermore possible that other factors, such as early polygyny, an early start to a surname's stock, or Y-chromosome linked fecundity might provide an additional boost to a single family's population.

## Appendix A: The Plant DNA results

Fuller details of the DNA results for Plant are given and are being updated elsewhere.[75] The full results are too extensive to include here and Table 2 gives just a brief sample of the Y-STR markers that have been measured. It corresponds to the worldwide dataset of independent volunteers in Figure 10(a). Some of the independent volunteers are omitted for the `English-based' dataset of the Discussion section and these are shown in Table 2 in red. Here, `English-based' means those whose ancestral line was known to have been in England as late as 1881. The remaining volunteers in Table 2 make up the dataset of `early emigrants' which have, for a significant time, experienced growth conditions different from those in England. For example, the number of Plants in the UK has grown by a factor of 2.01 since 1881 whereas those in the USA have grown (since 1880) by 3.36 though this might be due partly to further Plants emigrating from England.

In Table 2, PMH denotes the "Plant Modal Haplotype" of the matches; discrepancies from these values are highlighted in bold. Apart from PMH, a Plant project code for each volunteer is shown in the first column, with P denoting the spelling Plant and PT for Plantt. The column headings are the DYS labels of the first 12 Y-STR markers only.

| *DYS* | *393* | *390* | *19* | *391* | *385a* | *385b* | *426* | *388* | *439* | *389I* | *392* | *389II* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMH | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P1a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | **12** | 29 |
| P1b | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P1c | 13 | 24 | 14 | 11 | **12** | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P2a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 13 | 11 | 13 | 13 | 29 |
| P5a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |

[75] http://www.plant-fhg.org.uk/dna.html and http://www.familytreedna.com/public/plant/

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P6a | 13 | **23** | 14 | **10** | 11 | **15** | 12 | 12 | 11 | **14** | 13 | **30** |
| P7a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P7b | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P8a | 13 | 24 | 14 | 11 | 11 | **15** | 12 | 12 | **13** | 13 | 13 | **30** |
| P9a | 13 | 24 | 14 | 11 | **12** | **15** | 12 | 12 | **12** | **14** | 13 | **30** |
| P11a | 13 | 24 | 14 | 11 | 11 | **15** | 12 | 12 | **13** | **12** | 13 | **28** |
| P12a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P13a | 13 | **23** | **16** | 11 | 11 | **15** | 12 | 12 | **12** | 13 | 13 | 29 |
| P14a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P16a | 13 | 24 | **15** | 11 | 11 | 14 | 12 | 12 | **13** | 13 | 13 | 29 |
| P17a | 13 | 24 | 14 | 11 | 11 | **11** | 12 | 12 | 11 | 13 | 13 | 29 |
| P18a | 13 | 24 | 14 | 11 | **13** | **15** | 12 | 12 | **12** | **14** | 13 | **30** |
| P19a | 13 | 24 | 14 | 11 | **10** | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P20a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P21a | 13 | 24 | **15** | 11 | 11 | 14 | 12 | 12 | **13** | 13 | 13 | 29 |
| P22a | 13 | **25** | 14 | **10** | 11 | 14 | 12 | 12 | **12** | **14** | 13 | **30** |
| P23a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P24a | 13 | 24 | 14 | **10** | 11 | 14 | 12 | 12 | **12** | 13 | 13 | **30** |
| P25a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P26a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P27a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P28a | 13 | 24 | 14 | 11 | 11 | **13** | 12 | 12 | 11 | 13 | 13 | 29 |
| P29a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P30a | 13 | 24 | 14 | 11 | 11 | **13** | 12 | 12 | 11 | 13 | 13 | 29 |
| P31a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| P32a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| PT1a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| PT2a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |
| PT3a | 13 | 24 | 14 | 11 | 11 | 14 | 12 | 12 | 11 | 13 | 13 | 29 |

**Table 2: An extract of 12-marker Y-STR results for independent Plant volunteers.**

## Appendix B: The Simulation Model

In this appendix we describe the simulation model used to explore the probabilities of outcomes associated with unusual population growth. Our objective is to estimate the probability distribution associated with a surname line begun by a single progenitor in 1311 and continuing to 1881. We address this objective using Monte Carlo simulation. In this technique, the model includes a random variable whose values are generated according to a probability distribution. In our case, the random variable is the number of sons born to each family and surviving to reproduce in the next generation. The Monte Carlo simulation functions by simulating the dynamics of a single surname family a very large number of times (in our case, one million times) and taking statistics on the outcome of the simulations. These statistics are considered to provide an accurate estimate of the distribution of possible outcomes of the system being simulated (in our case, the growth of a surname population). The simulations of course depend on exactly which random numbers are used, and the purpose of having a large number of simulations is to ensure that this dependence is minimal. In our case, by using one million simulations we ensure that important quantities such as the size of the largest family do not change based on the specific set of random variables generated.

The model keeps track of reproducing males. The basic model assumes a 1:1 sex ratio and considers only males that survive to procreate in the next generation. The number of male children born to each father and surviving to adulthood is assumed to be a random variable drawn from a Poisson distribution. This is the most common way of selecting the size of a group of individuals, and is used in all simulation models of this type. For purposes of brevity we will not continue in the description of the model to specify that we only include male offspring who survive into adulthood; this will be implicit in the discussion. The Poisson distribution is characterized by a single parameter: the mean (in our case, the mean number of surviving male children in each family). This number is computed according to the theory of branching processes as described by Pinsky and Karlin[76] from the rate of population change in each generation. Population data were taken from Broadberry et al. for the period from 1311 to 1541, from Wrigley and Schofield[77] for the period from 1541 to 1801, and from census records for the period from 1801 to 1881.

In models that include early polygyny, we model polygyny by having each male in a generation in which polygyny is practised generate a number of male offspring equal to the sum of *n* Poisson distributed random variables, where *n* is the number of wives of the male, and has a fixed value for each generation in which polygyny exists.

The generation time is a key variable in the simulation. Although the human generation time is often taken to be about 25 years, recent research suggests that it is longer, possibly as long as 35 years.[78] We use a generation time of 30 years. Based on the assumption that the generation time can be taken to be the mean maternal age at birth, Wrigley and Schofield show that in England this age had a consistent value of about 31 to 32 years from the sixteenth through the nineteenth century. We

[76] M A Pinsky and S Karlin, *An Introduction to Stochastic Modeling* (Academic Press, Boston, 2011).

[77] E A Wrigley and R S Schofield, *The Population History of England 1541-1871: A Reconstruction* (Cambridge University Press, 1981).

[78] The following paper, for example, uses a paternal generation time of 30 years: B McEvoy and D G Bradley (2006) *Y-chromosomes and the extent of patrilineal ancestry in Irish surnames*, Hum Genet 119, 212-219. On the other hand, T E King, S J Ballereau, K E Schurer and M A Jobling (2006) *Genetic Signatures of Coancestry within Surnames*, Current Biology, 16, 384-388 adopt a value of 35 years.

took the value of 30 as a round number that reflects a possibly shorter generation time during earlier centuries.

## Appendix C: Family inheritance of genetic traits

A father's Y-chromosome is passed on to his sons essentially unchanged and so on down the generations. The human Y-chromosome molecule is smaller than the other chromosomes and it is currently reckoned to contain only 51 protein-coding genes and 119 non-(protein)-coding genes.[79] This is a tiny fraction of the genes in all of the 46 human chromosomes, which more generally are passed down through mixed male and female ancestral lines. It might hence be questioned how much a few proteins, coded by the Y-chromosome, could influence the shared characteristics of the males in a single-surname family. Each individual male might instead be expected to be dominantly influenced, for most of his characteristics, by autosomal genes, inherited from his preponderance of cognate lines coming through from the ancestral family trees of both his father and his mother.

However, the authors of SDFH[80] comment as follows.

*Bryan Sykes was impressed by the way his own surname had ramified over the centuries ... The name Sykes had, he wrote, `increased way above theoretical expectations' and this caused him to speculate what might lie behind that expansion. In the process, he outlined how the majority of the family names would naturally become extinct, over several generations in a stable population, and then went on to say how much more rapidly that would happen if the males in one of the families had a Y chromosome that somehow produced more sons. This offers a new perspective on the successful expansion of some distinctive surnames.*

In our simulations, we place more emphasis on fortuitous growth and the different historical growth factors of the overall population in different regions. As the above SDFH extract implies, the diverse rates for the dying out of families is a consideration since, to achieve the required overall growth, some families will need to grow faster to make up for those that become extinct.

For a male, his sex-determining Y-chromosome, inherited just purely down his male agnate line, is particularly associated with fertility in his biological germ-line, by which he passes on the coding in his Y-chromosome to his sons. We can hence conceive of a male-line family's consistent fertility,

[79] http://www.ensembl.org/Homo_sapiens/Location/Chromosome?chr=Y;r=Y:1-59373566 retrieved 3 Dec 2012.

[80] G.Redmonds, T.King and D.Hey, Surnames, DNA, and Family History, (Oxford University Press, 2011), p. 72.

brought about by their near identical Y-chromosome, and this might influence the growth of a single male-line family. This might arise, for example, if the inherited Y-chromosome leads to a small excess of fertile sons, above the norm for the more general population, and this particular mechanism has the advantage that it can be checked out in part by baptism and birth data. Such inherited fecundity for sons might augment the so-called random “genetic drift” whereby some male-line families, or family branches, grow disproportionately to others purely because of the random chance of having more fertile sons.

The available baptism data suggest that any historical imbalance in the male:female ratio, for the specific example of Plant offspring, is small. The method used[81] is probably not entirely accurate for this sex ratio though it allows a *like with like* comparison of different surnames. Though more males than females might be conceived, the recorded male baptism rate might be depressed by more male than female neo-natal deaths. Here, Plant is evidently associated with a particular Y-chromosome whereas the multi-origin surname Smith is characterised by a more general population. In Table 3,[82] taking the data at face value, 51.5% of Plants baptised in Cheshire are male and also 51.5% across the border in Staffordshire. These values have statistical uncertainties, in view of the small numbers counted, though they marginally exceed the value counted for Smith, which is 50.0%.

| | 1551 to 1600 | 1601 to 1650 | 1651 to 1700 | 1701 to 1750 |
|---|---|---|---|---|
| | | | | |
| Plant in Staffs | 13:17 | 29:27 | 85:82 | 340:320 |
| Plant in Cheshire | 9:8 | 44:37 | 81:79 | 210:200 |
| Smith in Cheshire | 134:130 | | 1148:1147 | |

**Table 3: Male:Female recorded baptism numbers in the main Plant homeland.**

For modern times, modern medicine has no doubt reduced neo-natal mortality rates. More modern results were obtained using the Ancestry dataset of births.[83] For recorded Plant births between 1950 and 1970, the male:female ratio has been counted to be 3209:2940; and, in the first quarter of 1960,

[81] Some forenames of indeterminate sex were ignored, such as Lesley and Leslie.

[82] Data taken from https://familysearch.org/ by counting male and female forenames, not including variant spellings of the surname. Dataset accessed 13th and 14th December 2012.

[83] http://search.ancestry.co.uk/search/category.aspx?cat=34 England & Wales, Birth Index, 1916-2005.

1422:1264 for Smith. In other words, the modern percentage of recorded male births was 52.2% for Plant and 53.0% for Smith. Again, the values cannot be taken as precise due to statistical uncertainties in the number counted. However, we may note that both of these values are close to a normally supposed value for England of 52.5% though variations of a few percent have been reported for general populations. This well recognized small excess of males is typically ignored on the grounds that males are likely to die younger than females. For the purposes of our simulations, we are mainly concerned with the possibility that a particular surname family might have a Y-chromosome that leads to an excess of males over the norm for other families.

Ideally, we would wish to know if an excess of sons over daughters led on to these sons surviving to become disproportionally many fertile fathers. These baptism and birth data do not include any information about the longer-term survivability and fecundity of the sons. Though not fully conclusive, these available results suggest that plausible effects, arising from a particular Y-chromosome, have not been demonstrated to be large. It has been the general case that child mortality for sons is higher than for daughters and, in discussing the results of our simulations, we have assumed a 1:1 ratio of males to females that survive.

Clearly, there are other factors besides male:female imbalance and pure chance that could lead to an abnormal number of sons fathering further sons in a family, such as lower mortality for the family due to favourable economic circumstances. There are also other molecular mechanisms, besides just those of Y-chromosome genes on natality, by which paternal identity might influence our findings: these are generally too speculative for explicit inclusion in our simulations, though they are perhaps worth mentioning in passing as a subject of topical interest.

The epigenetics revolution[84] of the past decade has identified other mechanisms by which paternal inheritance can influence future generations of offspring. One such mechanism is the chemical imprinting of autosomal genes with methyl markers that identify that this copy of the gene has come from the father as opposed to the mother. These markers in humans can regulate how much the particular DNA gene on any of 46 chromosomes is expressed to produce an associated protein. Another molecular mechanism is the production of small lengths of RNA (miRNA) transcribed from so-called “non-coding” genes on a chromosome. Thus, “non-coding” genes on the Y-chromosome might participate in regulating genes on other chromosomes, such as on the larger X-chromosome which is involved in brain function.

---

[84] Nessa Carey, *The Epigenetics Revolution: How Modern Biology is Rewriting Our Understanding of Genetics, Disease and Inheritance* (2012).

Though highly controversial, this mechanism of wider influences from the human Y-chromosome on paternally-inherited traits adds a possible twist to the spectre that those Plants who have volunteered for the Y-DNA test might not be typical of the whole surname. In other words, it is not impossible that there could have been a bias in the testing towards a particular family who were more genetically predisposed to take the DNA test than other Plants. In other words, this family might have had a particular Y-chromosome that led indirectly to a greater mental affinity towards volunteering for the test, because of an atypical family predisposition towards scientific enquiry for example. That might have led to an increased tendency for the sample of DNA tested Plants to display this particular family's Y-DNA matching result, disproportionately to those with various non-matching Y-DNA who might, in principle, be more characteristic of the Plant population as a whole, which might have descended from several different Plant origins with different genetic inclinations.

Such a bias in the Y-DNA data, though it seems rather unlikely, might even have conceivably affected most so-called "single origin" studies of surnames. In the Plant study, independent volunteers were DNA tested. Academic studies generally solicit a random selection of those with the surname throughout a large region, such as particular counties around West Yorkshire in England in the case of the initial Sykes study. There might conceivably be a bias towards those with a particular mental disposition towards accepting the invitation to participate in the study. This could lead to an over-estimate of the extent to which a populous surname is single origin. Epigenetics is still in its infancy, however, and any such Y-linked bias remains at least very highly speculative.

## Appendix D: Polygyny as a possible explanation of large families

Polygyny (i.e., a man taking many women) can help in explaining the high populations of some single male-line families. It has been proposed, for example, that polygyny explains the large population growths of some Irish patronymics associated with early medieval families.[85] Such an explanation holds for a single-surname family, however, only if the offspring inherit the true father's surname. For example, it has been objected that bastard Plants in England would not have inherited their father's surname and, hitherto, a polygyny argument has not been applied by others to the English surname Sykes. It is hence relevant to consider how the objection to paternal surname inheritance, for polygynous children, might be countered.

Two ways in which polygynous offspring could inherit the paternal surname are: (a) local custom allowed it; or, (b) infidelities were with women who already held the surname such that she passed on her own as well as the true father's surname to their children. These presume a concentration of

[85] T E King and M A Jobling (2009) *Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames*, Molecular Biology and Evolution, 26(5), pp. 1093-1102.

the surname in a locality. This, for example, might conceivably arise for the locative or topological surname Sykes meaning `from the', or `living near', the ditches.

In connection with large single familes, R.A.McKinley[86] comments, "in some parts of Britain, especially in the north of England, there are some areas where some locative surnames have ramified greatly. ... in such areas as south east Lancashire, the dales region of West Yorkshire, or the Yorkshire woollen district. ... The locative surnames which proliferated are nearly all ones which originated in the parts where they expanded, and most were derived from the names of small inhabited localities, either detached farms or hamlets. ... it must nevertheless be suspected that each surname began as the name of a single family ... even if the genealogical evidence to prove this does not exist". This single-family presumption for these surnames can be questioned, however. Thus, such a name as Sykes might conceivably have been ascribed to several unrelated families living at or near the eponymous location and just one of the families could have grown fortuitously, and also partly by polygyny with interloping men from the same fortuitous family which fed back into the surname.

With the advent of DNA testing, there has been a shift towards considering the rigour of a single origin contention for a populous surname. We may consider both of the aforesaid cases: (a) customary inheritance of the father's name for bastard; and, (b) the mother had the same name as the father. In both cases, a consequence of polygyny is that the infidelities would not show up in the Y-DNA evidence as false paternity events (NPEs) since there would be no break between the inherited surname and the inherited Y-chromosome. As a result, mismatching DNA signatures cannot be interpreted as representing a usual rate of NPEs, since most NPEs would not show up; they would more likely imply instead multi-family origins to the surname. Thus, where around half the tested men DNA match, as in the initial Sykes study, we might even go so far as supposing that the 50% DNA mismatches represent different origins and the normal rate of infidelity has instead doubled the remainder from 25% of the surname's population. This significantly reduces the deduced fraction of the surname population that is a single-origin family even before invoking that an abnormal extent of polygyny has contributed to this family becoming unusually large. More normally, it would be assumed that a usual rate of NPEs had reduced the extent of the DNA matching and this loss was balanced by Y-chromosome egressions to other surnames.

As well as for Sykes, we might consider the situation in more detail for Plant. The main concentration of Plants is found near the Welsh Marches and the Welsh meaning of *plant* is 'children'. Also, as in Ireland, there is evidence that polygyny was recognised in medieval Welsh Law. One might conceivably conjecture that the Plant name was initially coined for the 'many children' of a single family, or for earlier related family stock known in Welsh as "the children or the clan", or from the

[86] R.A.McKinley, *A History of British Surnames*, (London and New York, 1990), pp. 60-61.

place of such a clan. For example, there is 1310 reference to the manor of *la Planteland* in Monmouthshire and this name might refer to fertile land for raising plants and for breeding animals though not necessarily for the breeding of a clan of children. [87]

Turning more specifically to the possibility that the father's name was passed to bastards by local custom, it is often suggested that the origins of hereditary surnames in England related to the inheritance of office or property, inspired by the customs of the French nobility. Certainly the known early surnames were recorded in legal documents, often for the purpose of recording imposed fines, and they were perhaps ascribed to a peasant family by the authorities rather than because the peasants themselves had much, if any, say in their own surname. To this extent, it is relevant to consider the hereditary rights of bastards which were persistently a consideration for the legal authorities.

The historian, Rodulfus Glaber, was a monk at Cluny in eastern France, who died c. 1046. He approved of the Norman dukes and seems to have accepted the transmission of their office through 'concubines' which he defends by Old Testament precedence and that of the illegitimate birth of Constantine the Great.[88] If office could be transferred to bastards perhaps so could surnames. This is in general disagreement, however, with the Cannon Law of the church.

In Ireland, Brehon Law allows polygyny (albeit while citing the authority of the Old Testament) and other actions which Canon Law expressly forbid.[89] Brehon Law was effectively outlawed by the Statutes of Kilkenny in 1367 and the policy of Surrender and Regrant.

Welsh Law usually applied in the Welsh Marches as well as areas ruled by Welsh princes. Admittedly, the Welsh used a different naming convention from that of surnames but this illustrates an intermingling of customs. In a dispute, for example, between Gruffydd ap Gwenwynwyn and Roger Mortimer, Gruffydd wanted to apply English Law but, in 1281, the royal justices upheld Roger Mortimer's wish that Welsh Law should apply as the lands concerned lay in Wales. In Welsh property

---

[87] In the Callendar of Patent Rolls there is for 1310 Oct 10 Carmyle, Commission to John ap Adam, Master John Martel and John de Pateshulle to enquire ...; also to report on any defects in the castles of Struguyl and Turegi and manors of la Planteland, Tudenham and Berton, the weir, and the fencing of the park, what men Hugh le Despenser had placed on the works of the castle and manors, and what the works were.

[88] Rodulfi Glabri Hustorium libri quinque, tr. John France, Oxford Medieval Texts, Oxford, Clarendon Press, 1989, pp. lvii, 164-5, 204-5.

[89] D.A.Binchy, Introduction in *Corpus Iuris Hibernici*, p. ix.

Law, illegitimate sons were entitled to an equal share with the legitimate sons, provided they had been acknowledged by the father. This was the provision which differed most from Canon law. The recognition of polygyny in Wales may have been drawing to a close in the 13th century; but there was still recognition of the rights of the male offspring of such relationships. As the Iorwerth 13th century text puts it:

*The law of the church says that no-one is entitled to patrimony save the father's eldest son by his wedded wife. The law of Hywel adjudges it to the youngest son as to the eldest, and judges that the father's sin and his illegality should not be set against the son for his patrimony.*

Also relevant in this regard is the list preserved in several law books of nine sexual unions, *Naw Cynyweddi Deithiog*. The *Naw Cynyweddi* lists nine unions which seem unlikely to have met with Ecclesiastical approval. This is because some of them may plainly coexist with other unions in which either or both parties are involved. In other words, the list (similar to Irish lists) seems to presuppose a society which permitted polygyny. [90]

English and Cannon Law favoured primogeniture whereby the eldest legitimate son had sole rights to inheritance. However, even under English Law, no-one questions that others, besides the eldest legitimate son, could inherit a surname. Blackstone[91] states:

*Yet he [a bastard] may gain a surname by reputation though he has none by inheritance. All other children have a settlement in their father's parish; but a bastard in the parish where born, for he has no father.*

In short, if early polygyny was such that it allowed a surname to be passed on to polygynous children, it could contribute to a possible explanation of an abnormally large male-line family within a surname’s population. This could feature in the British Isles, as a supplement to less controversial explanations, in areas of mixed Celtic-Norman influence in particular.

[90] Huw Pryce (1993) *Native law and the church in medieval Wales*, p. 109.

[91] Blackstone's 18th century *Commentaries of the Laws of England*, Vol. I, ed. W. Morrison (London, 2001) pp. 352-53.