Getting the most from a surname study: semantics, DNA and computer modelling

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. 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.[1] In this paper, we consider the evidence that some frequent surnames could be single source; this would imply that a single family has grown abnormally large. For some populous surnames, they have a geographical distribution that might be thought to be consistent with a single origin though, 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. Instead, they are likely to maintain that statistical analyses of the data should be properly performed. For example, King and Jobling (2009)[2] considered forty English surnames and found no statistically significant correlation between the supposed semantic category of a surname and its degree of DNA matching into single male-line families. As a specific example that we here describe in some detail, little can be deduced about the inter-relatedness of those called Plant from the assumption of a semantic category, such as by arguing that it is locative and hence single-origin, or occupational and hence multi-origin. By comparison, more surely, we discuss the DNA evidence that this name's main family grew unusually. Though motivated initially by the evidence of unusual growth for Plant, we extend our deliberations more generally to other surnames.

Guided by the empirical evidence, our computer simulations identify various reasons for a surname family's prolific growth. In particular, chance is a main factor, along with favourable conditions during the Industrial Age when overall population growth took off, evidently earlier in some regions than in others. Also, the modelling suggests that some additional factor such as polygyny or resilience to plague or favourable economic circumstance, after an early start to a hereditary surname, is beneficial in seeing a family through initial precarious times, sustaining its survival

[1] 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.

through to a small but real chance of subsequent proliferation in favourable Industrial Age conditions.

Contents

Introduction

In the past decade, DNA testing has prompted revisions to our understanding of surnames. In some cases, this extends to their supposed meanings.[3]

Meaning has long featured strongly in surname studies, with linguists freely accepting such a semantic device as metonymy when postulating a surname's meaning, sometimes to the neglect of other possibilities. Sometimes, a name's meaning has been used to suggest a hypothesis for its origins, though, for example, a fuller consideration allows that there are at least four different semantic hypotheses for the Plant surname's origins and growth. Some people have taken surname meanings to indicate the likely number of geographical origins to a name: an `occupational name', for example, could have many origins unless, of course, it derives from a particular family's rare adoption of its seminal word(s).

Further clues as to whether a surname might be single- or multi-origin can be gleaned from modern mapping techniques. As supporting evidence, DNA investigations can help to ascertain how far a single family has spread. With the advent of flourishing DNA studies of surnames, there is renewed interest in whether chosen English surnames are single-, plural-, or multi-origin, indicating one, a few, or many distinct origins. A received wisdom is that populous surnames in England are multi-origin[4] and a limited DNA study by King and Jobling (K&J) goes some way to support that.[5] However, following a pioneering DNA study of the Sykes surname, over a decade ago, there formed a contrary view that at least *some* common surnames are single- rather than multi-origin.

The geographical distribution of a surname provides a relatively ready criterion for deciding whether a surname might be single-origin. Ideally, this distribution should be traced back to the origins of each surname, perhaps as far as 800 years ago. However, the data for that early are generally incomplete and unreliable. The available studies have been used inductively to generalise that a surname's distribution often remained 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.

Surnames that are both populous and single-origin are not the norm amongst surnames and so we consider them as atypical arising from exceptional circumstance. Where possible, in due course, we extend our approach to the relative mathematical probabilities of unlikely events, so that we can assess various explanations for the occurrence of some populous single-origin surnames.

DNA and some surname distribution evidence

Some surname distribution maps from Steve Archer's 1881 *Surname Atlas* CD are used, in this section, as a guide to infer some early places of likely origin for prolific surnames. For less populous surnames, the chance migrations of a few individuals are more likely to affect a surname's overall geographical distribution. A note of caution is needed even for populous surnames. If their initial

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

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

[5] 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.

population was small, in medieval times, the chance migration of a few could conceivably have affected the name's overall subsequent distribution, such as by 1881, though, as a simple template, we follow others in assuming, at least initially, that peasants in medieval times were generally immobile. Such an assumption can be thought to hold true in most cases.

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.

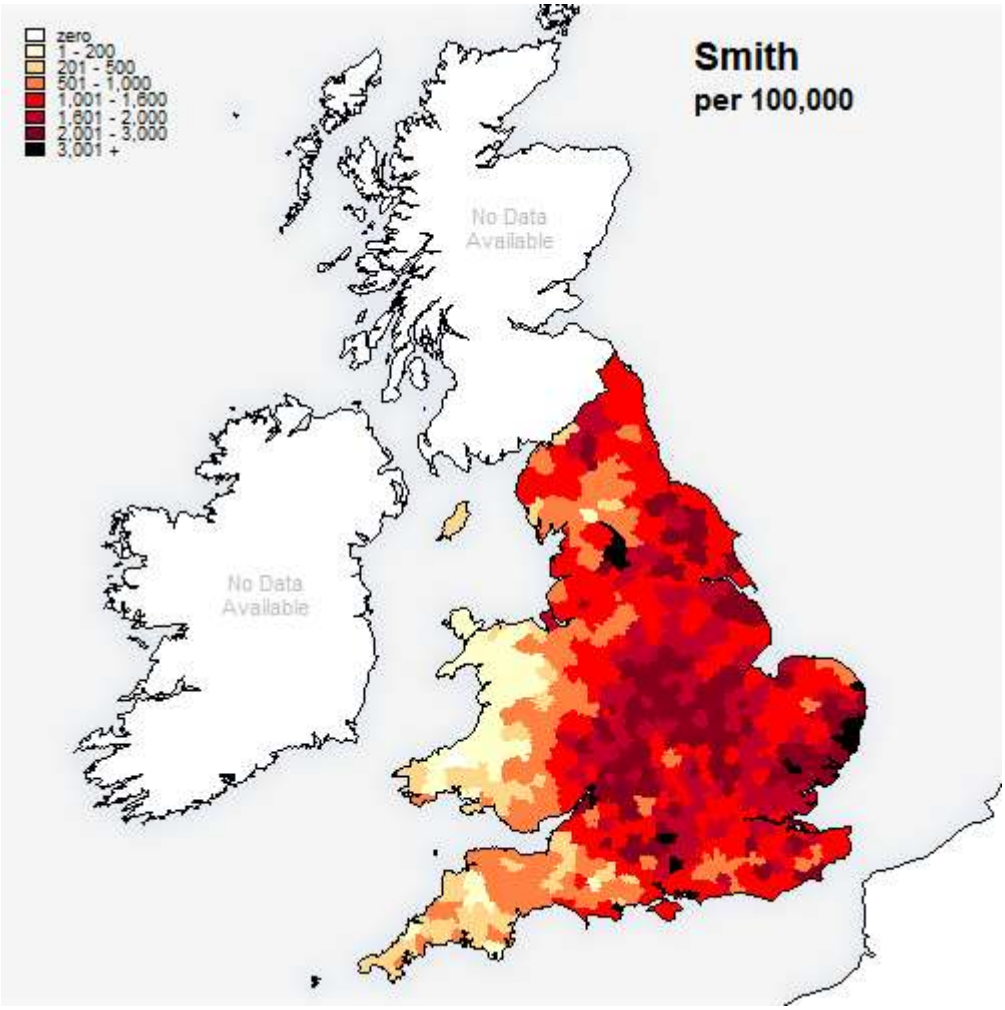


Figure 1: The distribution of the surname Smith in England and Wales in 1881. The shading represents the number of Smiths per 100,000 people, for each Poor Law Union.

The largest cluster of matching Y-STR signatures for Smith, in the K&J study, was not a compact one; put another way, this diffuse cluster formed a loosely-connected pattern when just differences in

the DNA signatures of participating Smiths were plotted out without regard to their geographical locality. The whole cluster amounted to 15% of the total number of the Smith men who were DNA tested. With its diffuse nature, this DNA cluster could be explained largely by the groupings of Y-STR signatures in the R1b1 haplogroup, which is common in the general UK population; in other words, this cluster can be regarded as dating back to pre-surname times since when many mutations have occurred leading to only loose matching around a modal signature in the DNA results. K&J accordingly concluded that there was no overwhelming evidence, for Smith, for the occurrence of particular single families that were closely related in the past 800 years since the formation of this very common surname.

The less clear-cut situation for Metcalf

As a different example, Redmonds et al have suggested that the surname Metcalf is "single-origin".[6] As pointed out by Debbie Kennett,[7] the DNA results for this surname show six distinct DNA clusters, which can be associated with 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 the DNA results. 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.

(a) (b)

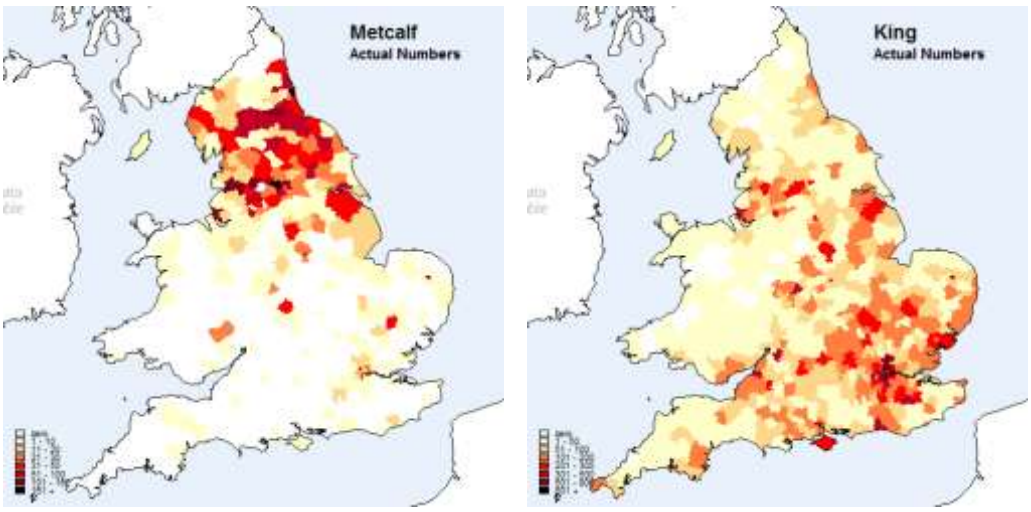


Figure 2: The distribution of the two populous surnames, (a) Metcalf and (b) King, in 1881. The shading here represents the actual number bearing the surname in each Poor Law Union.

According to the Surname Dictionary of Reaney and Wilson,[8] Metcalf is thought to have derived as a nickname meaning someone who resembles a calf to be fattened for meat. The variant Medcalf is

[6] George Redmonds, Turi King and David Hey, *Surnames, DNA & Family History*, (Oxford, 2011), pp. 73-4.
[7] Debbie Kennett, (Jan-Mar 2012), Book Review of Surnames, DNA & Family History, Journal of One-Name Studies, Vol. 11, Issue 1, p. 29.
[8] P H Reaney and R M Wilson, *A Dictionary of British Surnames*, 2nd Edition (1976).

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.

Available DNA evidence in connection with some other populous surnames

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, the UK population in 1881 was 65,233 and this was widely distributed (Figure 2(b)). In the K&J study, the 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.

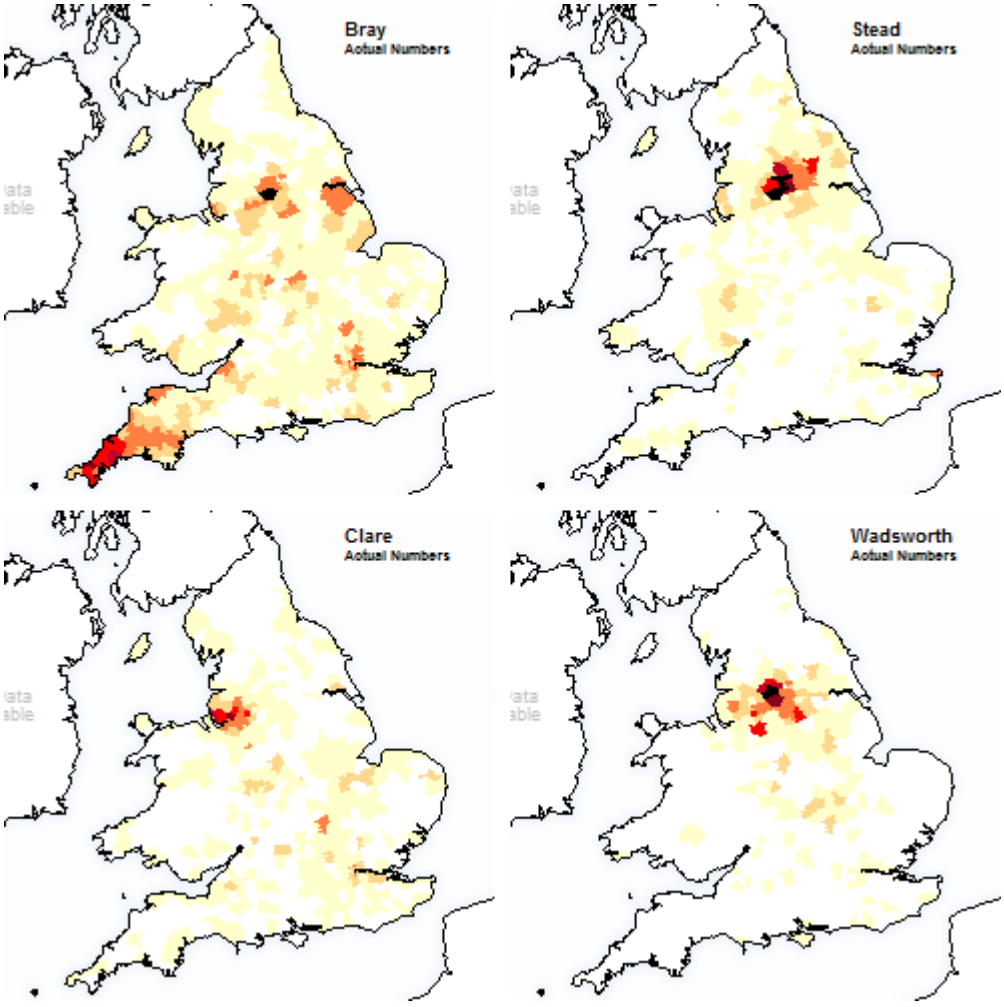


Figure 3: Distributions in 1881 of some populous surnames from the K&J study.

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.

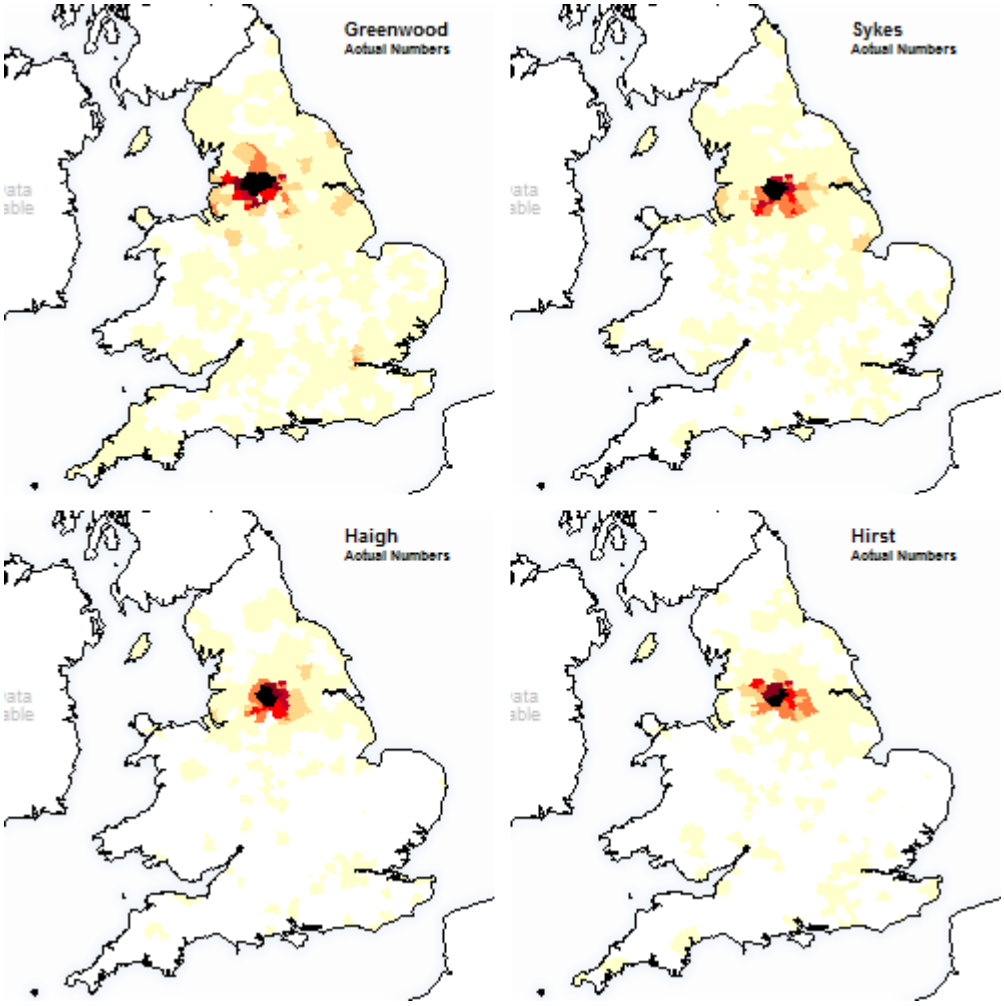


Figure 4: The four most populous single-origin contenders in the West Riding of Yorkshire.

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. Though Shaw is more populous than these,[9] it is omitted from Figure 4

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

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 based solely on their 1881 distributions; that supposition has not been confirmed by DNA evidence, apart from some supporting evidence for Sykes.

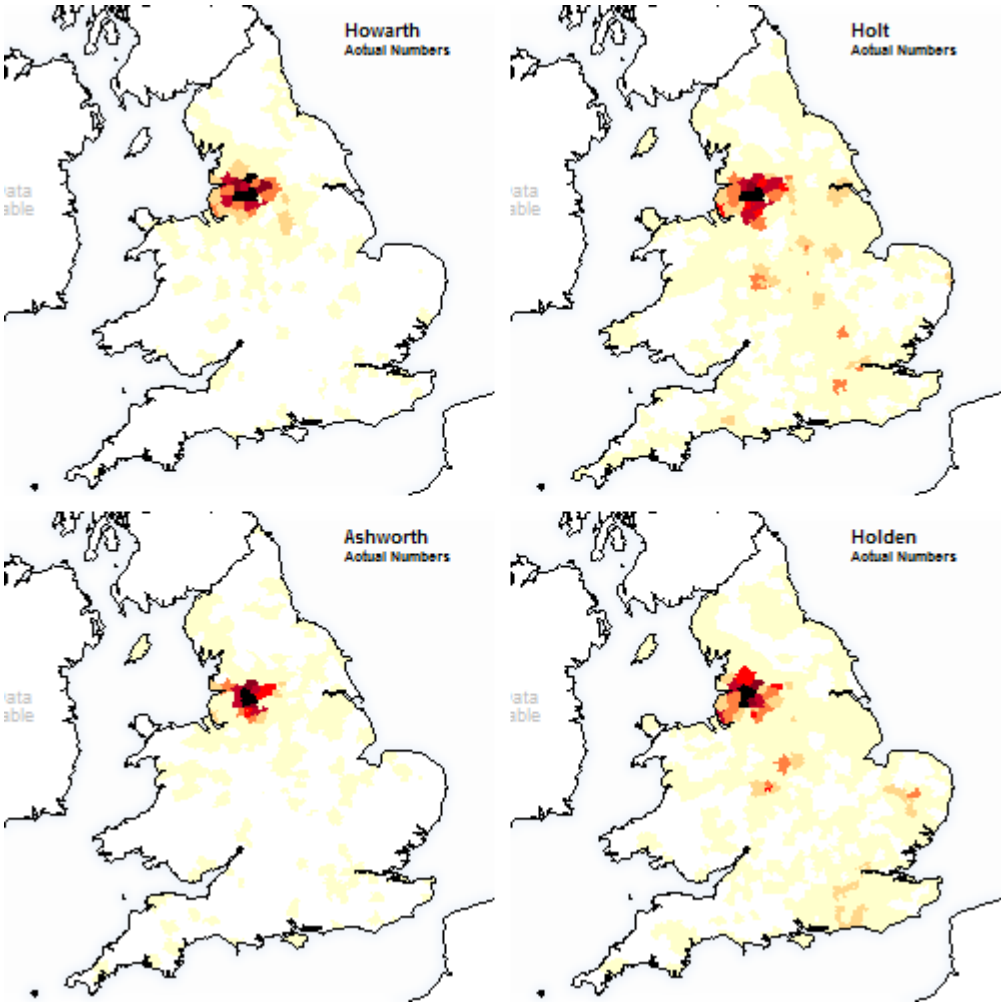


Figure 5: The four most populous single-ancestor contenders in Lancashire.

George Redmonds has taken the early DNA study for Sykes[10] as confirmation that some populous surnames in West Yorkshire are single-origin.[11] The Sykes DNA study was low resolution and measured only four Y-STR markers. It is none-the-less remarkable that these Sykes surname results displayed a single cluster made up of DNA matches for 43% of the tested men, though statistically this percentage has a 7% uncertainty when generalised to the whole surname.[12] This somewhat surprising DNA result marked a beginning to renewed interest, a little over a decade ago, in whether

[10] B Sykes and C Irven (2000) *Surnames and the Y Chromosome*, Am. J. Hum. Gen. 66(4), pp. 1417-1419.
[11] George Redmonds, *Names and History* (2004) pp. 26-31.
[12] 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

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 a large number of large genetic clusters, including the one found in the early English study, but this American study was both non-random and relied on relatively-few emigrants to America[13] and hence is unsuitable for our purposes.

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.[14] 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).

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 Warpentake of the English counties, between 1761 and 1841.[15] 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 cannot be ruled out, however, as a significant factor in the population growths in specific Hundreds, with new work opportunities perhaps attracting many to these particular areas of high growths in overall population.

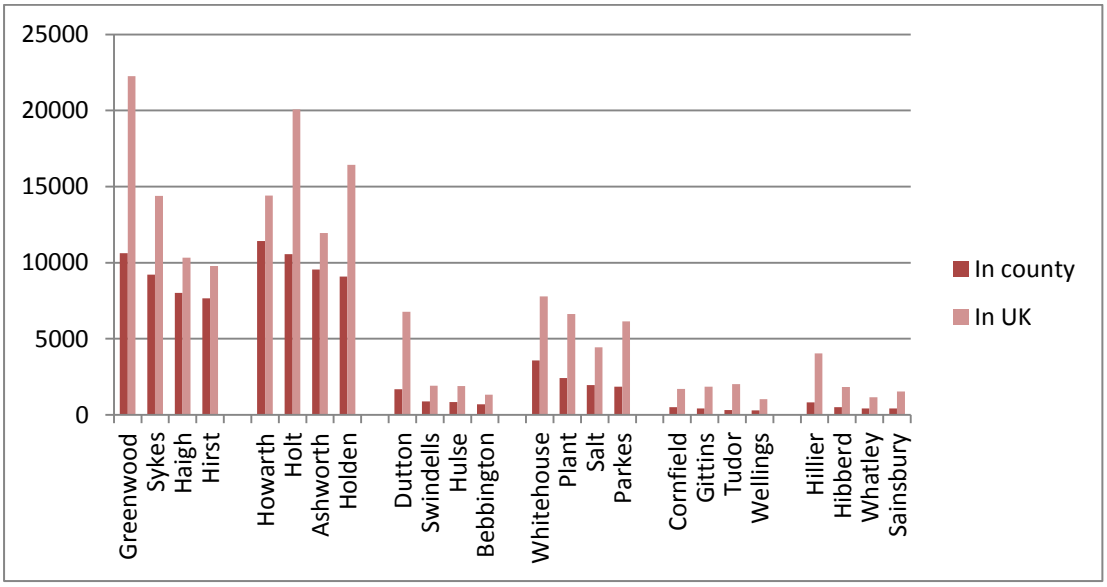


Figure 6: Populations in 1881 of the four largest single-ancestor contenders in each of six counties: West Yorkshire; Lancashire; Cheshire; Staffordshire; Shropshire; and Wiltshire.

[13] See the Chapter *DNA and Surnames* in *The Surname Handbook: A Guide to Family Name Research in the 21st Century* by Debbie Kennett (to be published).

[14] 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).

[15] 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 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 to the right of Figure 6.

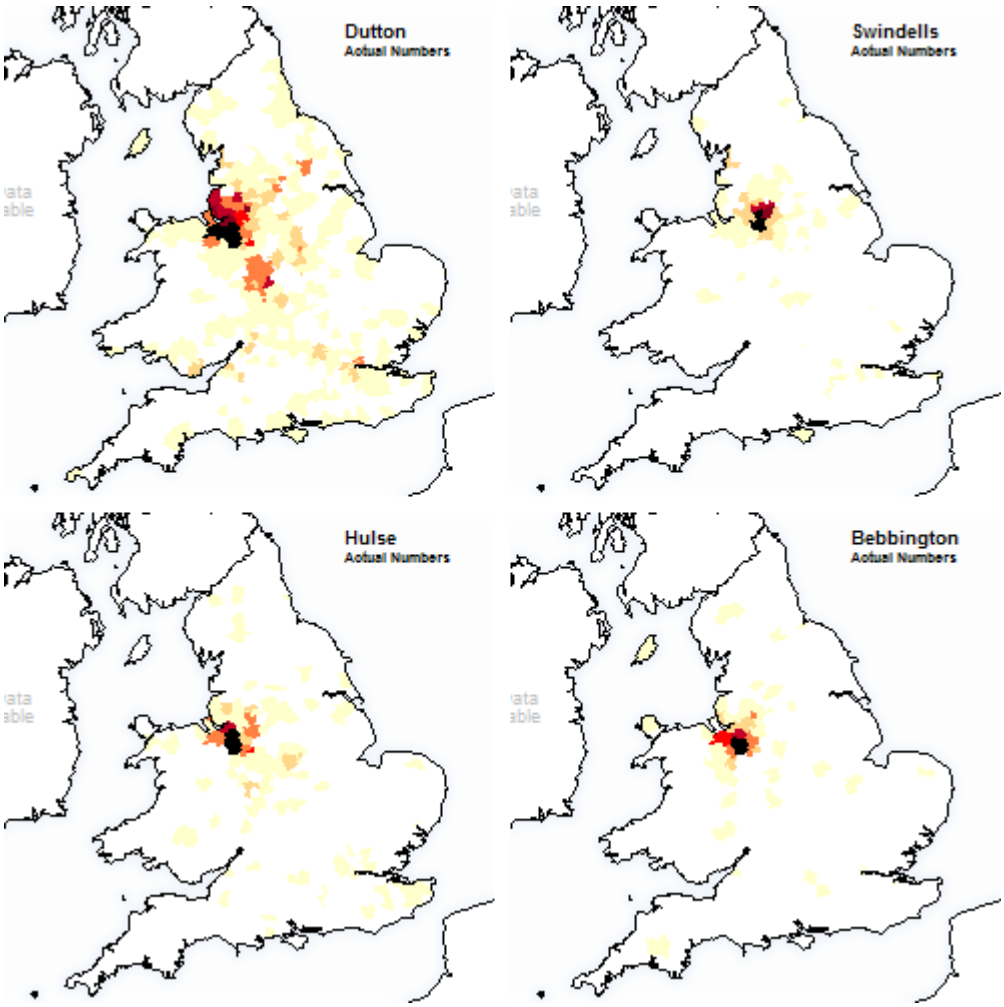


Figure 7: The four most populous single-origin contenders in Cheshire.

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 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.[16]

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%).

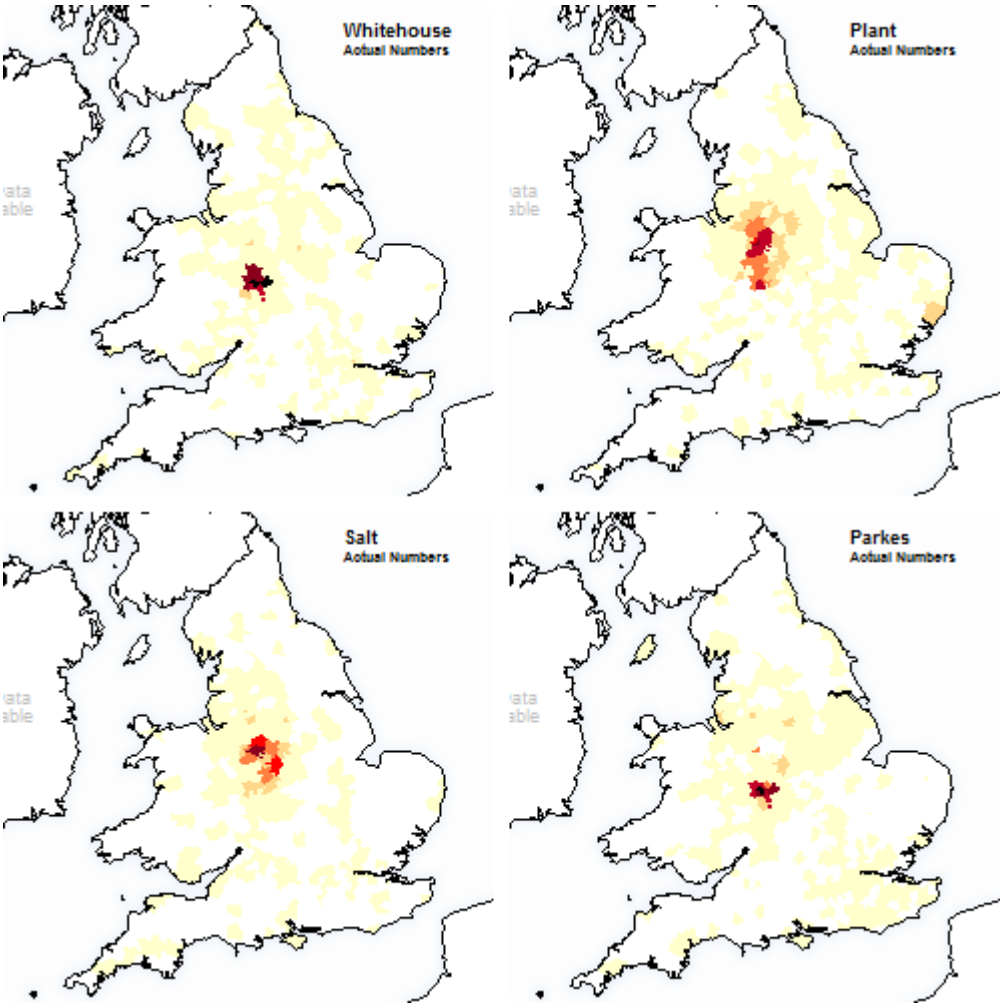


Figure 8: The four most populous single-source surname contenders in Staffordshire

[16] 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).

Populous single-origin contenders in Cheshire and Staffordshire

The single-origin contenders that are included[17] 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 debatable as a convincing 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 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.[18] None of these is contained solely in Staffordshire, though their distributions show a concentration there.

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 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 as possible single-origin contenders. Many of the most populous contenders appear to be in 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.

Taking another county, with moderately populous single-origin contenders, the most populous surname in Staffordshire that contends 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.

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

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

[18] 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.

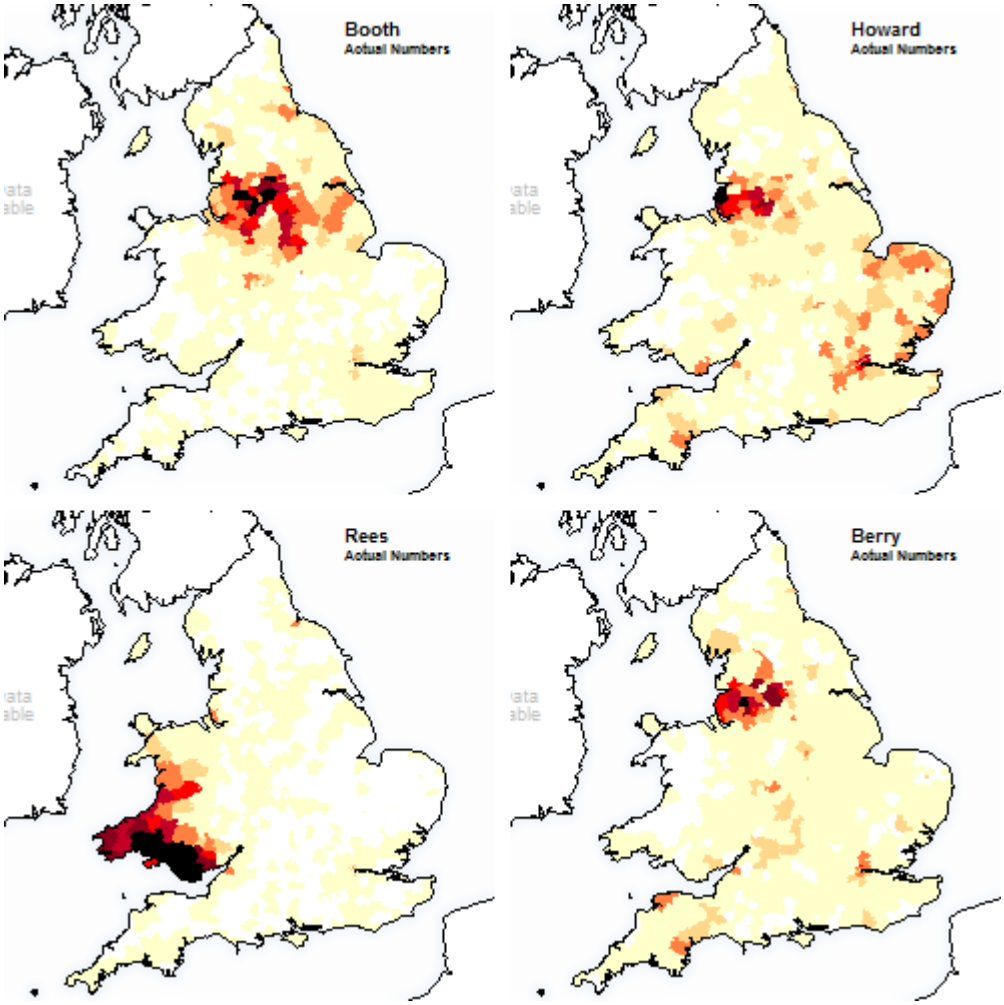


Figure 9: Less likely, though seemingly the most feasible, single-origin contenders amongst the 145 most populous surnames in the UK in 1881

Considering a surname in more detail: Plant

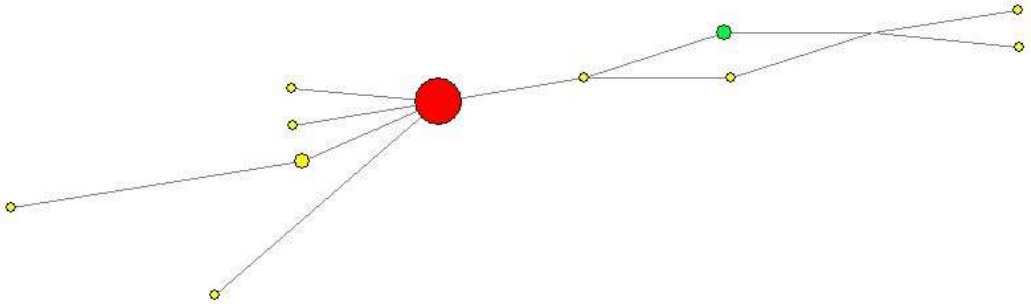
As we have a wide range of detailed information for Plant, we shall consider this populous Staffordshire surname more fully.

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. The Plants whose Y-STR results are included in Figure 10 volunteered independently, except in the case of a pair who both thought their earliest known male-line ancestor had lived in Leicestershire. Some care has been taken in this DNA study to consider a random sample of the Plant surname.[19] Academic studies have used different procedures and conventional `snail' mail in trying to guarantee a random sample of a surname; the internet has been exploited for the Plant study.

[19] 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

(a)



(b)

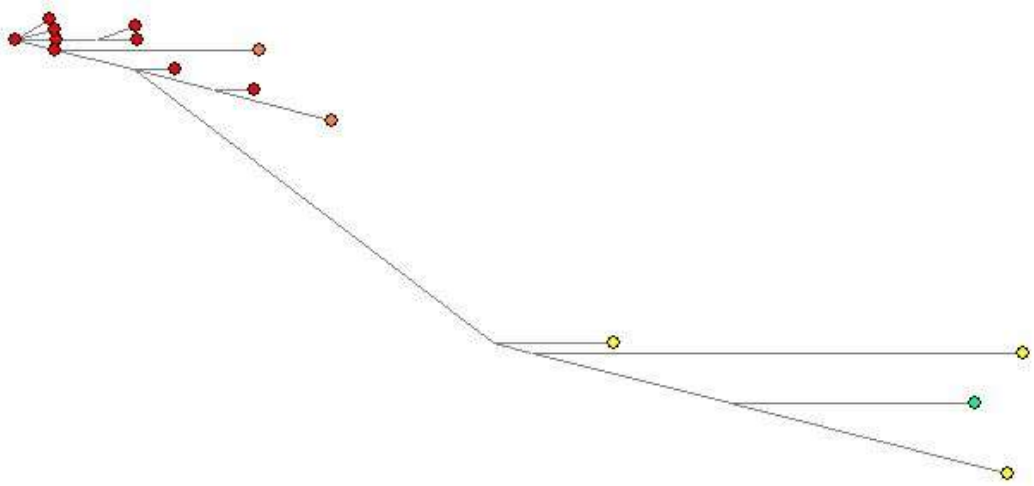


Figure 10: DNA Network diagram for Plant: (a) 9-marker diagram; (b) 37-marker diagram

The green circle in Figure 10(a) is associated with two participants with separate known ancestries, both in south Lincolnshire and, though no genealogical link has been found between them in the

documentary evidence, 25-marker DNA testing confirms that they have a shared male-line ancestor. Only one of them, who has had 37 Y-STR markers measured, is included in Figure 10(b)

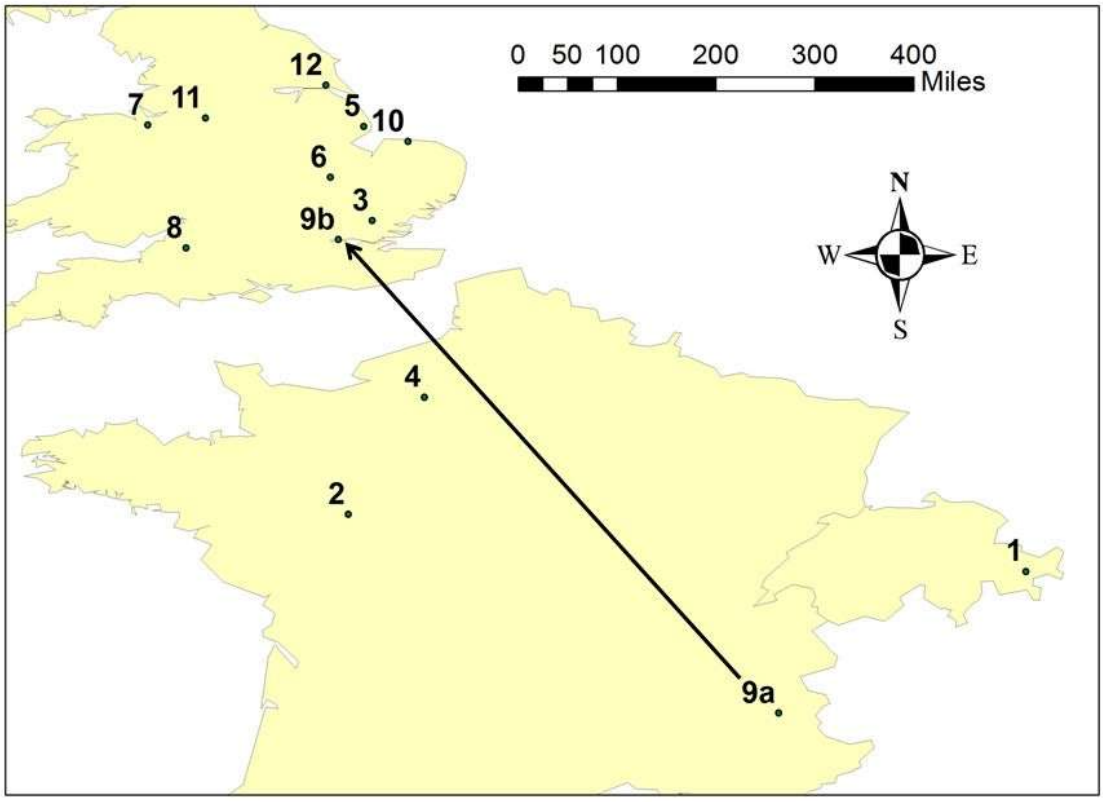
Otherwise, there is only one DNA cluster (red circles in Figure 10) yet found amongst the independent volunteers with the Plant surname. Their shared male-line ancestor, indicated by the DNA matching, might well date back to an early bearer of the Plant surname, perhaps located near the "main" English Plant family homeland, or even earlier, perhaps elsewhere, as is discussed later. The large red circle in Figure 10(a) represents a number of exactly matching Y-STR signatures at the 9-marker level which is replaced in Figure 10(b) by a cluster of near-matching small red circles. The reason for non-exact matching of the red circles in Figure 10(b) 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 outliers in 10(b), which are within the large red circle in 10(a), are coloured orange and they can be considered to be *possible* matches to the cluster of red circles in 10(b). A Deep Clade test has been carried out on one of the core in 10(b) of small red circles and this has determined that it has a sub-clade that is quite rare in England (currently estimated at only around 6% of the population). A Deep Clade test is now being carried out on the orange circles to see if they belong to the same sub-clade, hence confirming that they adequately match the main Plant family represented by the red circles.

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, there is an explanatory proviso 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-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.

The green circles might correspond to an entirely separate origin for the Plant surname or, like each yellow circle, they might correspond to an NPE from the same surname origin as that for the red circles. Taking the latter case, in the 9-marker data there are at present 20 matches (red circles in Figure 10(a)) and 11 non-matches (green and yellow circles in Figure 10(a)). Two subjects who did not volunteer independently are excluded. Considering only independent volunteers, 35.5% can be taken to be *observed* non-matches and this can be compared with *theoretical* expectation under the assumption that all non-matches are NPEs. If the false paternity rate per generation is p, there can be *expected* to 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=0.0189, to wit an NPE rate of 1.89% per generation which compares well with a "reasonable estimate" of 2% as used by K&J in their computations. In other words, it is perfectly feasible, within the realms of theoretical expectation, that surviving Plants descend from a single source for the name with the 35.5% of mismatches being due to NPEs. If this same calculation is performed using the 37 marker data of Figure 10(b), and excluding the two questionable subjects, the result is an NPE probability of 0.022. If the two questionable subjects are included, the NPE probability drops to 0.013.

None of the Plant volunteers has a Y-signature corresponding to the modal R1b1 signature of the UK general population. The available Plant results of those for whom only 9 comparable markers have been measured are not far distant from this most frequent result for the general population of Europe and North America. However, the Plant results are very distant from this modal signature

when more markers are compared. 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) and who 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 DNA evidence for another origin to the name, in contradistinction to many of the surnames studied by K&J.



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 modal Y-STR signature for Plant, at a 38-marker level, is given elsewhere.[20] One Plant volunteer in the main DNA cluster has been found to belong to the sub-clade R-P312+.[21] 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.[22] At that time, the term "single-origin" was usually reserved for relatively rare surnames that were almost entirely in a small local geographical area.[23] 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 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 "single ancestor".[24] By this, it is meant that the DNA-tested *living* Plants can all 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 those living bearers of the surname who have been DNA tested.

The DNA results for the living Plants might not relate to all of the first Plants. There might, for example, 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 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. We shall return to this question in the subsequent modelling.

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.[25] 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 proof that all early instances of the name belonged to the same Plant family.

---

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

[21] At the time of writing, this sub-clade is denoted R1b1b2a1b* and none of its currently known sub-clades has tested positive (M65-, M153-, U152-, L21-, L176.2-).

[22] 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.

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

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

[25] 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.

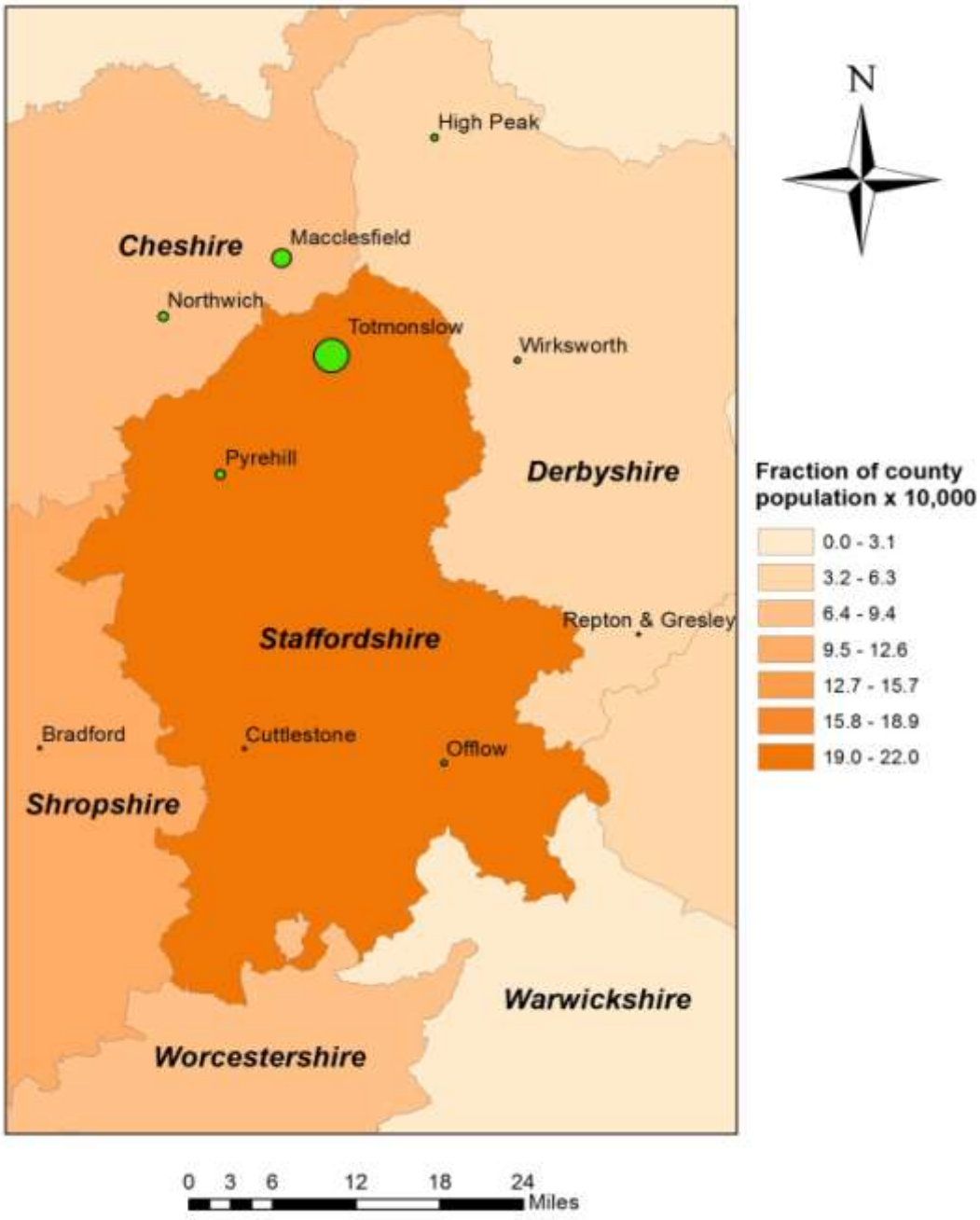


Figure 12: A cluster of Plant Hearth-Tax households (1662-89) is represented by the green circles. The graded brown represents the number of Plants per 10,000 in each county, in 1881

Some early instances of the Plant name could have been by-names that were not inherited; and, even initially inherited lines can be expected to have died out, entirely or largely. 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, if the same Plant family was not already more widely spread before it migrated to its main homeland.

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 Figure 12. We refer to this 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 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 either separate medieval origins to the name that died out or the early locations of a mobile family.

Some considerations pertaining to early surname migration

Though the general evidence for medieval England indicates that most families remained in their small local area, the widespread ramification in England of a few "single-origin" surnames might have begun by as early as the fourteenth century. At that time, not everyone was a villein, tied by law to their lord's local land.[26] For example, 10% rising to 20% of the population lived in a town.[27] 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 aristocracy, as well as the secular lords, were 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 to some recorded early Plants.

It is not inconceivable that early Plants travelled far and that their name dated back to very early times. It has even been claimed[28] 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, though it is relatively credible that the late medieval name Planta *might have* morphed to Plante and then Plant.

[26] Ian Mortimer, *The Time Traveller's Guide to Medieval England* (London 2009), pp.38-47.
[27] Christopher Dyer, *Making a living in the Middle Ages: the people of Britain 850-1520* (New Haven and London, 2002).
[28] G.R.de Beer (1952) Notes and Records of the Royal Society of London, p.8.

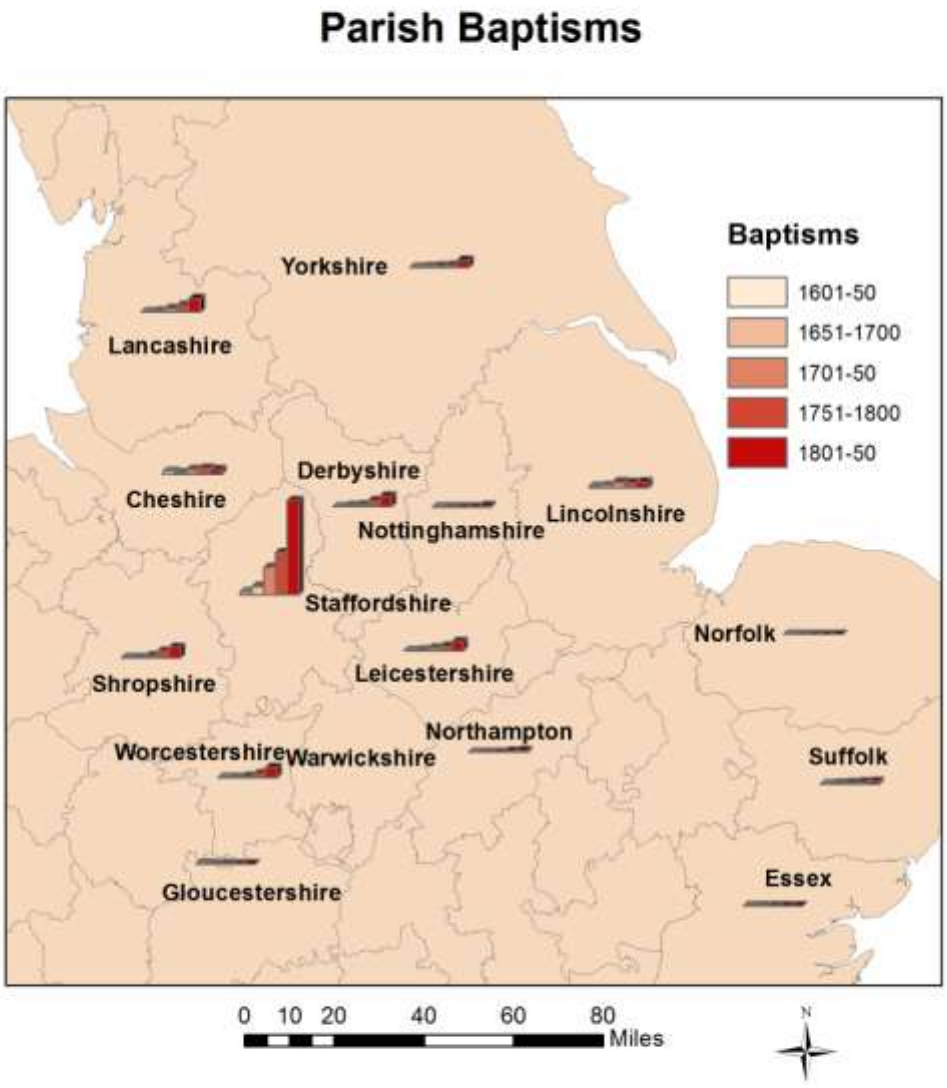


Figure 13: Distribution of Plant baptism records throughout several English counties (1601-1850)

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,[29] from 1601 to

[29] This data has been taken from the 1984 microfiche of the International Genealogical Index.

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.

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 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 less 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.

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 circle.

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 labelled 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 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).

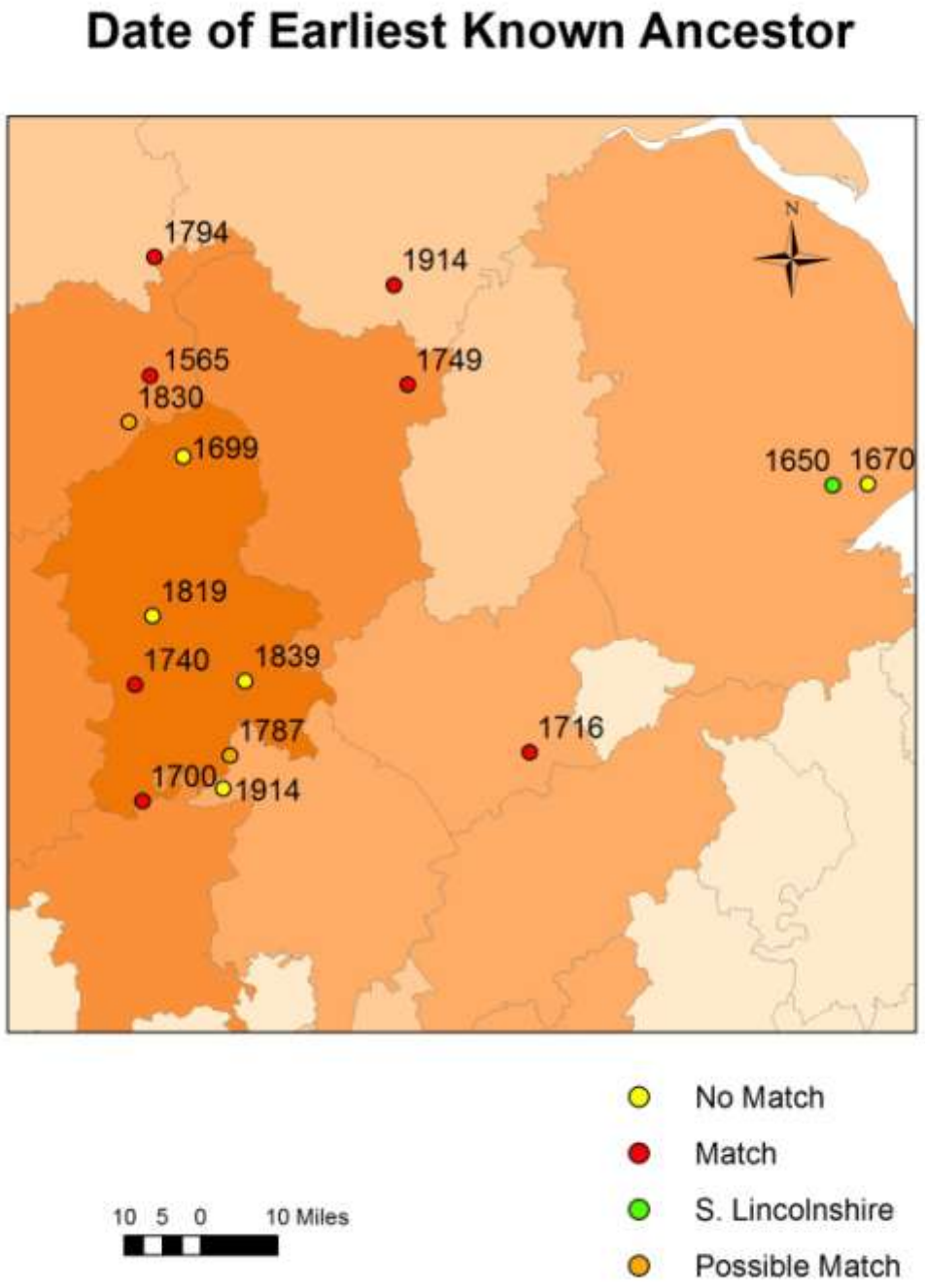


Figure 14: Distribution of DNA matching and mismatching ancestral Plants. Where two matching Plants have been found in the same local region only one circle is here included, labelled with the date of the earliest known matching Plant ancestor there.

Attempts to trace back the earliest known ancestors of those 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 Plant homeland (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 multi-origin contribution to the 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). 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 the English Plant family.

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*. As we shall show later below, it is not necessary however 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 but it is not clear that this was an origin to the hereditary surname or contributed much, if at all, to 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 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 family in the local geographical area. Relaxing the "single origin" contention, these could have supplied a few origins to Plant as a hereditary surname, some of whom 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,[30] *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,[31] suggesting sole as an ontological foundation to man's existence or soul. Certainly, the word *planta* means `to beget children' in Welsh in which *plant* means `children'.

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]'[32] 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'.[33] 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.[34]

Finally, 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.[35] Two such possibilities are as follows.

- 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, which can be related to a supposed attraction of a relatively late semantic development of the word *plant*.

The Oxford English Dictionary[36] 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 – the verb *to plant* had meanings `to found' or `to establish' from early times and these can be related on to industrious generation. We might wonder if such a meaning could have appealed to tradesmen around the Potteries of North Staffordshire, for example, perhaps enough to

[30] Robert Claiborne, *The Roots of English: A Handbook of Word Origins* (1989).
[31] 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.
[32] 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.
[33] Huntingdon manuscript HM 143, *Piers Plowman*, Passus XIX, ll. 9, 25-26, 101-102.
[34] http://www.plant-fhg.org.uk/soul.html
[35] George Redmonds, *Surnames and Genealogy: A New Approach* (Boston, 1997).
[36] OED, *plant* n(1) 6a.

cause 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 *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'.[37]

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

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[38] can be summarised as follows:

- 1360 Rand Plont(t) and Willo Plont(t) fined for stray animals in the Forrest
- 1363 Thom Plont(t) indicted
- 1370s Seven Plonts are mentioned in 35 separate entries for stray animals and pannage
- 1374 Thom Plontt had failed to pay the fine for pasturing a bullock in the 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

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.

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. 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[39] at Midgley on the Cheshire-Staffordshire county boundary (item dated 1374 in the above list), at the heart of the main Plant homeland.

[37] Nellie Kirkham, *Derbyshire Lead Mining through the Centuries* (Truro, 1968), p. 88.
[38] http://www.plant-fhg.org.uk/EarlyPlontsMacclesfield.pdf
[39] Clarice Stella Davies (1976) *A History of Macclesfield*, 12.

Though contentious, another hypothesis has been that the name was displaced to here from Warren lands. 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 - a transient Welsh influence on the Plant name might relate to this as a temporary location. 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 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 imagine that the parent haplogroup of its ancestry underwent a mass migration, around 4500 to 4300 years ago, up the river Danube through central Europe into Western Europe,[40] where the Y-chromosome of the ancestors of the main Plant family underwent a further SNP mutation (P312 also known as S116).[41] A nomenclature for this R-P312+ sub-clade has recently been revised to R1b1a2a1a1b* and it 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 ancestry might eventually tie down more specific locations, with relevant timescales, more tightly.

Returning to the documentary evidence of the past millennium, we can surmise that it is not impossible that the main Plant family could have carried on the name Planta from the Continent in the twelfth century, perhaps culturally if not genetically, though, as a less tenuous hypothesis, the main Plant family's name could have originated independently in or near fourteenth-century

[40] 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.

[41] Testing shows that all of the currently known 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.

Macclesfield Forest (position 11 in Figure 11) for individuals that much earlier had had Continental ancestry.

Modelling prolific population growth

To summarize some discussion from the previous sections, we propose 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. We shall also, in due course, consider the possibility of a multi-origin component to the name or an earlier origin. Irrespective of uncertainties about earlier times, the available DNA evidence suggests that a single fourteenth-century family had, by 1881, grown to a population in England of approximately 6,615 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.

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,[42] 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 be 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 6,500.

Two categories of explanations

There are a number of possible alternative explanations for the observation that the Plant surname has a much higher population than might be expected from the descendants of a single male ancestor in the fourteenth century. These explanations may be divided into two categories.

- The first category contends that Plant is actually a multi-origin surname, and that the indication of the DNA test results that it is single-origin is incorrect. Assuming that those selected for the DNA test do derive from a single origin, this explanation holds that it is only by chance that all of these test subjects descend from this origin and not from the many other origins that can be imagined to have existed.
- The second category of explanations posits that all Plants currently alive are indeed descended from the same individual (apart from the NPEs); and, for some reason, the descendants of this person experienced an unusual growth in number.

We recognize that these are extreme polar opposites, and that the true situation may be somewhere in between: to wit, that *some but not all* of the DNA mismatches result from NPEs and that *many but not all* of the Plants alive today are descended (apart from NPEs) from one person. As

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

such, our categories represent bounds on the range of possibilities. Since it is impossible with the present data to determine which of these alternatives is correct, we can only attempt to compare their probabilities. We have developed models to estimate the probabilities associated with various explanations for the size the Plant surname population. The estimation of probabilities associated with the second category, the single-origin hypothesis, is carried out using stochastic modelling. The estimation of probabilities associated with the first category, the multi-origin hypothesis, can be accomplished using much simpler coin-tossing models. We will deal with this first.

Prior to taking up the models, however, we must establish a standard that can be used to compare them. Statistical modelling is most commonly concerned with average behaviour or likely events, but we are concerned here with exceptional behaviour and unlikely events.

- If Plant is a multi-origin name, then the unlikely event is that the DNA results for the selected volunteers have remained consistent with those for a single origin surname. This is unlikely in so far as it contends that the majority of the men with this surname so far selected *just happen* to belong to the same family - by calling it the "same family" we here keep in mind that there is almost no pattern among those who do not DNA match and that these can be ascribed to the same family through NPEs.
- If Plant is a single-origin surname, then the unlikely event is that it has grown to such a large population.

To establish a standard of comparison between these two unlikely events, we will establish a "one in a hundred" rule, that is, we will determine the model properties needed in order that the given event has one chance in one hundred of occurring. We are acting under the assumption that an unlikely event has occurred and, although an event with a chance of one in one hundred is unlikely, it is not impossible.

Now we turn to the modelling.

First category of explanations

The first objective is to create a probabilistic model for the event that there are multiple Plant families but that, by chance, all of the individuals whose DNA was tested happen to belong to the same family, which we will call the "main" Plant family. Because of non-paternity events (NPEs), some individuals will not be male-line descendants of the original male Plant ancestor. These Plant individuals are not, however, male-line descendants from another male Plant origin but rather simply descendants of random male ancestors who have `infiltrated' the "main" Plant family. For our purposes, the male offspring of these infiltrators can be treated as part of the "main" Plant family; their number can be expected to be roughly balanced by the offspring of Plant males who have `infiltrated' other surnames and hence neglected for numerical purposes.

We start our reasoning by considering the possibility that there is more than one Plant family, from separate origins, besides the "main" one, but that by chance all of the individuals tested happen to belong to the "main" Plant family. Clearly the probability of this occurrence depends on the size of the "main" family relative to that of the other families from other Plant origins. It is this relationship we here calculate.

There is a very large population, which we can consider as essentially unlimited, of individuals available for DNA testing. None of the individuals tested were aware in advance of the testing of a particular familial relationship, so we can assume that each was tested independently. Therefore we can model the selection for DNA testing as a coin tossing procedure, with a coin that is not necessarily fair. We assume that if the coin lands heads, then the tested individual happens to have come from the "main" Plant family. If the coin lands tails, then the tested individual is not a member of this family.

There are 15 individuals in the 37-loci data set. Therefore we model the testing procedure as tossing the coin 15 times. The event that all individuals tested are members of the "main" family is that, in this model, the coin lands heads in each of the 15 tosses. If the two populations – that is, the "main" Plant family, as against the branches from all the other Plant origins combined – are of equal size, then the probability of this "15 heads" occurrence is vanishingly small: on the order of one in a million. In order for the probability to reach as high as one percent, the "main" Plant family must comprise approximately 74 percent of the total Plant population, and the families from all other Plant origins approximately 26 percent. This comes very close to Plant being an "effectively" single-origin family for the practical purposes of modelling the DNA results.

Second category of explanations

Now we consider the second category of explanations, that Plant is a single-origin surname that experienced unusually large growth.

Our simulation model is very similar to that used by Sturges and Hagget[43] and later by King and Jobling.[44] The limitations of these models are well known: they do not allow for emigration or immigration, nor do they allow for changing of surnames. 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 the Appendix. It was programmed in the R computer language, and the code is available from the authors. The simulation only includes males, so it represents about one half of the total population. It tracks the number of male descendants of a population of males 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 every 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 the Appendix for justification). The year 1311 is the year in the early fourteenth century that differs from 1881 by an even multiple of 30.

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 members will not be actively reproducing because they are too young, and a

[43] C M Sturges and B C Hagget, *Inheritance of English Surnames* (London, 1987).
[44] 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.

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 use the value one half. Our model ignores emigration, so that it considers the entire descendant population to be in England in 1881. The total male population in that census with the surname Plant is approximately 3,300. We initially take the reproducing population to be one half of this, or 1,650. Therefore the objective of our simulation is to determine the conditions in the model under which there are 1,650 active male descendants in 1881 of a single active male progenitor in 1311.

The simulations begin with an initial set of 500,000 individuals, which is an estimate of the number of reproducing males in England in 1311. The model keeps track of the number of males in the lines of descent of each of these individuals. We call the male-line descendants from each of the 500,000 progenitors a single "surname family". Earlier family relationships between these progenitors are ignored. We also here ignore the effects of NPEs, since male-line egressions from the surname to different surnames can be expected to be balanced by introgressions into the surname from other surnames and we are here only concerned with the number of the progenitor's active male progeny.

Null model computations

We consider several alternatives. The first, which we call the "null model," is that there is no inherent difference in any of the lines of descent. This implies that a large population of descendants in 1881 of any one progenitor in 1311 is due only to chance. In each generation some families will have more surviving sons than others purely due to chance, and the null model represents the situation in which this element of chance is the only difference between lines of descent.

Of the 500,000 single "surname families" present in 1311, a total of 40,987, or approximately 8.2 percent, survive to 1881. Figure 15 shows a histogram of the distribution of the population sizes of the surviving "families" of descent.

Though it can hardly be seen in Figure 15, the most populous surname family contains 579 individuals. Therefore we can conclude that it is virtually impossible that paternal lines stemming from a single male in 1311, which only differed due to random chance, could grow to a size of 1,650 reproducing males in 1881. If the Plant lines did experience this growth then this family must be doubly exceptional. It must have some property different from the average among English families; and, among families that have this reproductive advantage, it must have enjoyed unusually large growth due to random chance.

The 25-marker DNA test results provided an earlier indication that 41% of the subjects belonged to the "main" group. If we consider the extreme assumption that there were *no* NPEs in the history of the surname, then this would indicate that, of the approximately 1,650 active Plant males in 1881, approximately 41%, or about 675, were descended from the progenitor. This still indicates that Plant is an extreme surname, even under these most conservative assumptions.

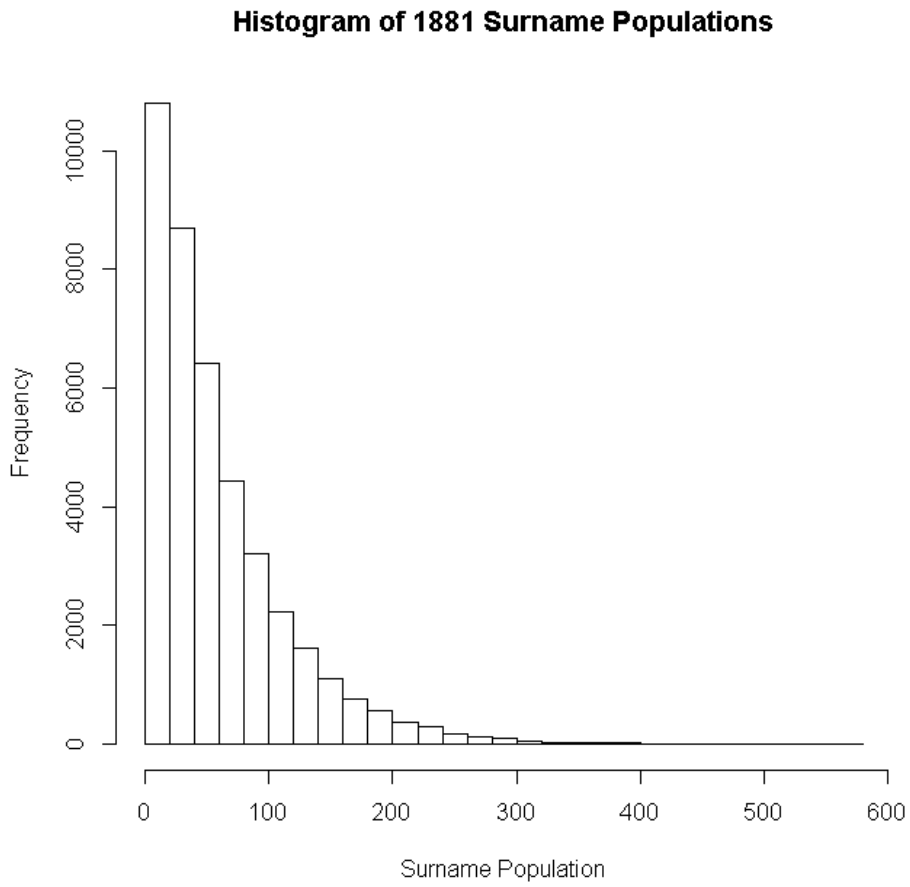


Figure 15: Histogram of single "surname family" populations by 1881, under the null model

Three alternative further models

We consider three alternatives to the "null model."

- 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.
- The third is an "enhancement" model, in which a family's growth rate is enhanced during some or all of the generations of the simulation.

Consistent with our standard for comparison of the alternate hypotheses, we ran each simulation repeatedly, adjusting the parameters until one percent of the surviving families had a population of at least 1,650.

First we consider the simulations of the "early polygyny" model. 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 in order to generate a simulation in which one percent of the 500,000 families have an 1881 value of at least 1,650 requires 12 wives per male for two generations. It cannot be achieved in one generation of polygyny even with 50 wives. These values are sufficiently extreme that it seems unlikely that early polygyny alone could account for the large size of the family.

Next we consider the "early start" model. In order that one percent of the surname families in the simulation have a population of at least 1,650, a population size of 175 in 1311 was required. Starting with a single individual and assuming a growth rate consistent with that used in the rest of the simulation, fifteen generations (i.e., 450 years) would be required to reach a population size of approximately 175. Although this is not impossible, there is little credence that Plant was a fully hereditary surname by the mid-ninth century.

Finally, we consider the "enhancement" model. Here we assume that one particular family enjoys an enhanced reproductive rate during some or all of the nineteen generations of the model up to 1881.

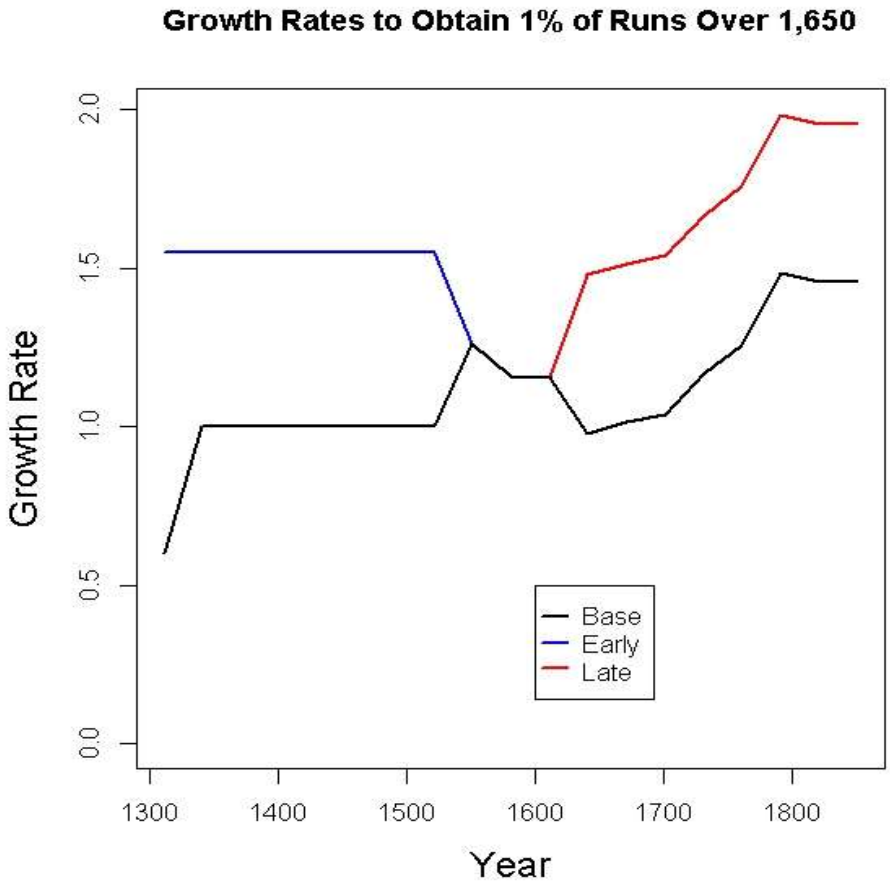


Figure 16: The base growth rates used in the models and the higher values used for early or late enhancement of the rates

The early and late enhancement sub-models

We consider two sub-models of the "enhancement model". The first is "early enhancement", in which the family's reproductive rate is enhanced for the first few generations. This might simulate, for example, a reduced effect of the plague known as the Black Death or advantages due to some aspect of the geographical location or social status of the family. The second sub-model is "late enhancement", in which the reproductive rate is enhanced over some of the last steps. This might reflect a differential effect on the family of the Civil War and/or the Industrial Revolution.

Figure 16 shows the base growth rate of the models (black line) and also the enhanced growth rates required (coloured lines) for one percent of the simulated lines to achieve a population of 1,650 in 1881. The base growth rate is deduced from available data for the total English population at various dates. Perhaps the most striking feature of the simulations is that about the same degree of early enhancement (blue line) as of late enhancement (red line) is required to achieve the requisite large population. In other words, there is no advantage to an early enhancement over a late enhancement. The primary advantage conveyed by early polygyny or early enhancement is simply to increase the chance of the surname surviving into the period of higher growth rates, when surnames have a chance of surviving and growing.

Comparing the model results

Figure 17 shows population sizes of individual runs of the four primary alternative models, each of which reach values in 1881 of approximately 1,650. The "early enhancement" (brown line) and "early start" (green line) models both display exceptional growth prior to 1700 by when they both coincide with the "early polygyny" (magenta line) model. Although its growth rate enhancement begins in 1671, the "late enhancement" (orange line) model does not catch up to the others until the end of the nineteenth century.

Of the three alternatives to the null model – early polygyny, early start, and early or late enhancement – only the late enhancement model seems by itself at all capable of explaining with realistic parameter values a family size as large as that of the Plant surname in 1881. Even this model requires an enhancement in population growth that seems rather difficult to accept as realistic.

The simulations do, however, provide an indication of a combination of these effects that may indicate an explanation, namely, some level of early polygyny (or early start or early enhancement) together with an enhanced growth rate during the Industrial Revolution. The blue line in Figure 17 corresponds to a combination of early polygyny and late growth-rate enhancement. Here, one percent of the surname families reach the population of 1,650 in a simulation in which there are initially ten polygynous wives and in which the growth rate during the Industrial Revolution is enhanced by a value of 0.3.

The explanation for this can be seen by comparing the fraction of surviving families in 1671 between the null model and an early polygyny model. For the null model, only about eight percent of the families survive until 1671, while forty-three percent of the families in the early polygyny model with ten wives survive until this date. In other words, the advantage conveyed to a surname family by early polygyny, or early start, or early enhancement, is not so much to increase its eventual population size as rather to increase its probability of survival into the Industrial Revolution, when circumstances arise that engender high growth rates.

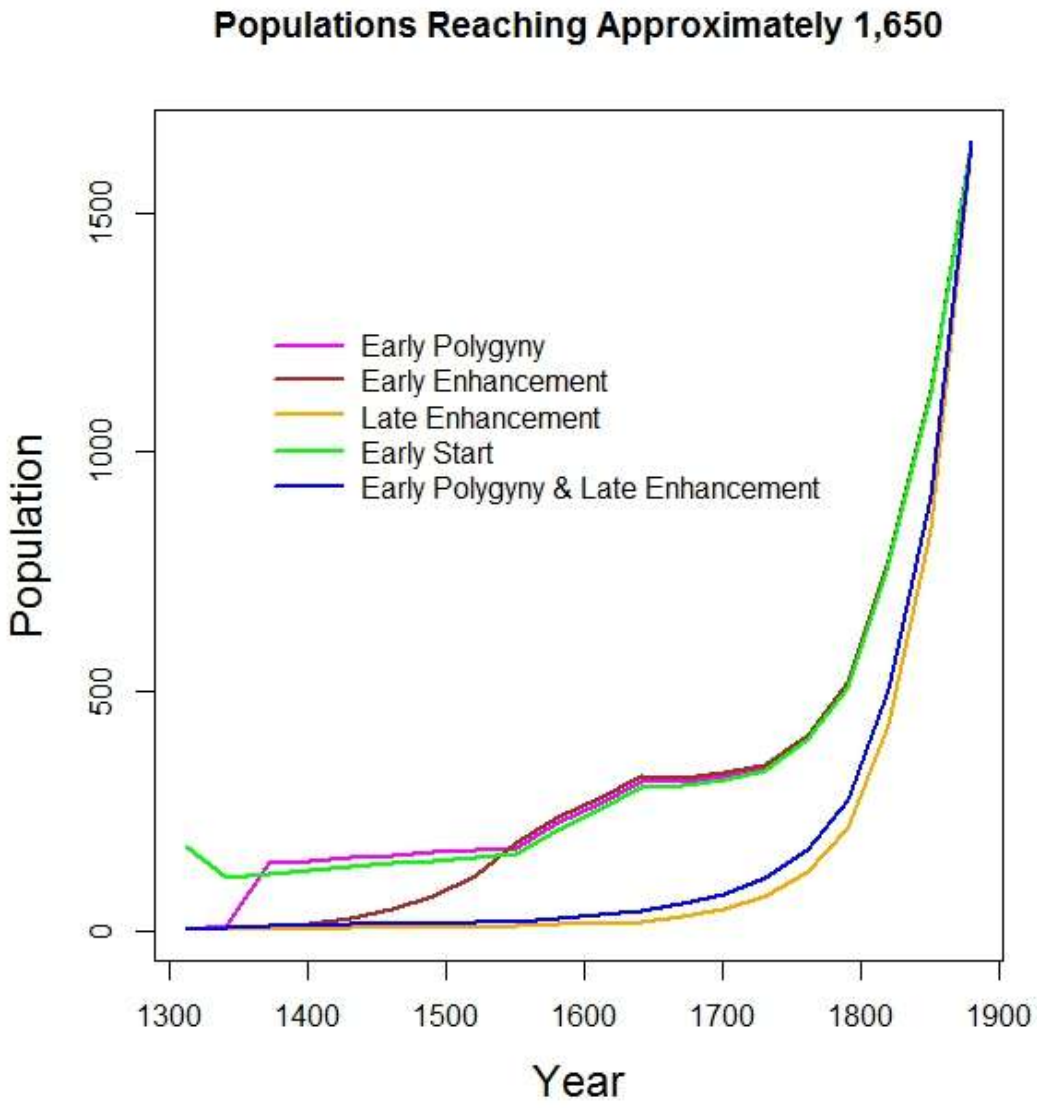


Figure 17: Population growth in various models for a surname family reaching a population size of 1650 in 1881

Conclusions

Most populous UK surnames appear to be multi-origin. However, it seems that the UK distribution of a surname such as Berry, which numbers as many as 23,700 in the 1881 UK Census, might conceivably be taken to have stemmed mostly from a single origin, on the basis of its 1881 distribution. More rigorous DNA studies might well help to clarify the picture beyond inferences based solely on a surname's geographical distribution, though the number of suitable DNA studies is as yet limited. Amateur DNA studies typically are not careful to consider a random sample of a surname's whole population, often seeking instead to check out particular family links and, if the surname is populous, this is often for a non-random selection from the surname.

Around the turn of the millennium, the early DNA evidence for Sykes, with its 1881 UK population of over 14,300, caused a stir when it was suggested that the living bearers of this populous surname were evidently from a "single origin". According to Sykes and Irven (2000), 43% of the living bearers of the Sykes surname around their main homeland DNA matched. This led to a controversial revision of the informed consensus that surnames that are this populous might not necessarily need to be considered to be multi-origin in order to explain their high populations.

Like Sykes, many of the most populous "single-origin contenders" are in West Yorkshire and Lancashire. They can be regarded as "contenders" in as much as they have a similarly compact geographical spread in 1881 as Sykes (Figure 4), with its 1881 UK population of 14,383. More cautiously, in terms of the higher resolution studies of King and Jobling (2009), it can be noted that Wadsworth is similarly geographically compact and has a cluster of 35% DNA matches despite an 1881 UK population of 4,175 (Figure 3). Considering a county adjacent to the south of Lancashire, the most populous contenders in Cheshire have smaller populations, as typified by the surname Swindells with its 1881 UK population of under 2,000. Moving yet another county further south, the most populous contenders that are concentrated in Staffordshire are rather more populous than in Cheshire: e.g., Whitehouse with its 1881 UK population of nearly 7,800 and Plant with over 6,600. Though Plant had a fourteenth-century geographical concentration around the border between Cheshire and Staffordshire, it mostly migrated somewhat southwards into Staffordshire. DNA testing has shown it to have a cluster of 64.5% DNA matches at the 9-marker level.

Considering in more detail a surname such as Plant, it is clear that caution is needed with the term "single origin". This surname may, or may not, have originated a few times, even if the descendants from only one origin have come to dominate its DNA results. Regardless of whether or not it was entirely "single origin", it is reasonable to propose that a single Plant family had grown to an 1881 UK population approaching 6,600.

We have been led to conclude that Plant is an unusual surname. This has in turn led us to attempt to quantify how unusual it is, and we have therefore attempted to determine properties that would make it a "one in a hundred" surname. That means, more or less, that we have tried to establish properties that the name must have in order to have a one percent chance of occurring. It is important to recognize, however, that because an event is unusual it does not mean that it is impossible. Indeed, an event that has a one percent chance of occurring will in fact tend to occur roughly one percent of the time, no matter how unusual it may seem when it actually does occur.

The results of the DNA testing for Plant, for example the matching of 65% of the samples at the 9-marker level and between 60% and 73% at the 37-marker level, with no substantial pattern among the non-matches, are consistent with the speculation that everyone tested had descended from the same individual, and that the non-matches are due to NPEs. It is very unlikely that such a selection of those tested from the Plant population will occur unless the vast majority of men with the surname Plant are descended from the "most reproduced" progenitor. The one in a hundred standard set in the paper was attained if the "main" family comprised 74% of the population. In order for there to be, for example, a 50% (instead of 1%) chance of selecting all test subjects from the same family, that family must be taken to comprise about 94% of the living Plant population. There is, of course, the possibility that some, though not all, of the tested individuals descend from other Plants, but the

probability that there have been few or no NPE's during the approximately twenty-three generations since the establishment of the surname is also very small.

By comparison, our initial computer simulations (Figures 16 and 17) suggest that the *one in a hundred* chance that a given single family originating in the fourteenth century would grow as large as this, as evident for a single Plant family, would happen only under favourable conditions. This can be expected to happen only rarely. However, this *theoretical* finding is in keeping with the *empirical* finding that at least one populous `effectively single origin' surname exists.

Various models have been considered in which one per cent of the progenitors in the early fourteenth century succeed in fathering a single family that grows as large as 6,600. The null model computations suggest that only around eight per cent of the families survive. Most of the families in the simulation that die out do so during the first few generations, when growth rates are only slightly above, or less than, one. However, favourable factors for a particular family, such as early polygyny, or early population growth enhancement, can greatly increase a surname family's early chance of survival. The initial population of a family is generally small and any required favourable factors need apply to only this small family's few individual members. However, our initial computations suggest that, taken one at a time, these early factors need to be surprisingly large in order to lead on, in themselves, to a sufficiently large 1881 family population. Extra growth in the times of the Industrial Revolution, when the general population was growing faster, can more readily provide an addition boost that has the requisite effect. However, we then need to consider that this extra boost has to apply to the whole of the surname's now-larger population.

A late boost to a family's population seems feasible if the surname is reasonably concentrated in a region where favourable conditions apply. We might also think of widespread land rights, or family-networked skills that are consistent with above-average living conditions and hence more prolific survival. Could "good genes" play a role? There are relatively few genes on the Y-chromosome which passes down male lines of a surname. Several deleterious inherited traits, however, are known to be sex related, and it is not impossible that some advantageous traits could also be. The possibility that a surname might carry favourable genes has been proposed by some as a possible explanation of the large size of some families. Here, favourable genes would imply above average fecundity, perhaps due to an increased tendency to survive childhood through to an early marriage followed by reproduction into relatively old age.

Some large male-line families might have had more than one surname. However, many of the most populous surnames appear to be multi-origin. Though not confirmed, it seems possible that the largest single-origin contenders might have 1881 UK populations of over 22,000 (Figures 4 and 9) implying around 5,500 reproducing males.

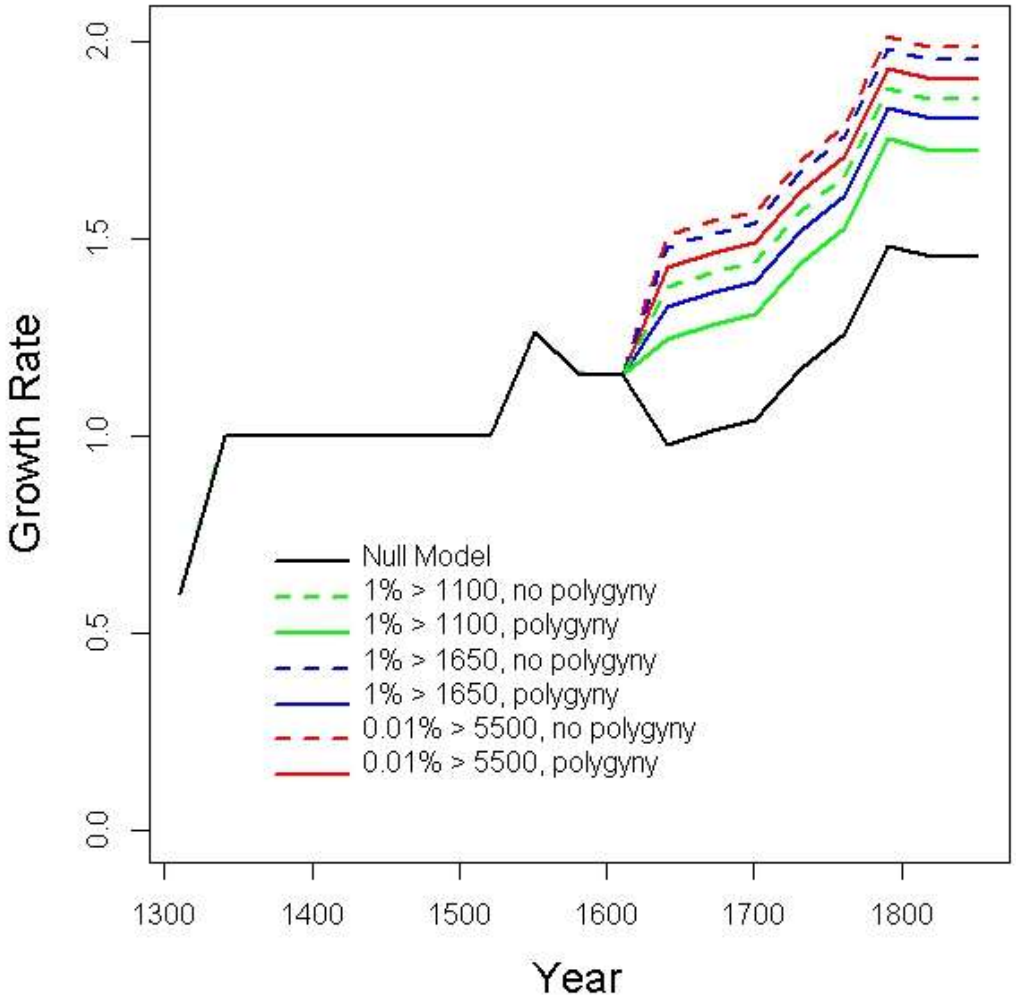


Figure 18: Enhanced growth rates required to reach various chances of particular family populations

The red late-enhancement lines in Figure 18 indicate the enhanced rates required to allow 0.01% of favoured families to grow to 5,500 reproducing males. Of each pair with a particular colour, the broken line is for no other favourable factor whereas the solid provides that there is also early polygyny involving 10 wives in the first generation. The red lines might be considered appropriate to the populous "single origin contenders" in West Yorkshire and Lancashire (Figures 4 and 5) where the Industrial Revolution came early and forcefully. There is rather less evidence to support such extreme late enhancement around North Staffordshire though the Pottery Industries thrived here early. However, for the Plant name, Welsh customs might have applied in its early main homeland around the northern border of Staffordshire allowing a supposition of early polygyny and/or there could have been some other early contributory factor such as favourable living conditions around

the Black Prince's vaccary.[45] The lowest late-enhancement lines might hence apply to Plant, with the solid green line applying if, for example, the number of Plant males in 1881 was *three times* the reproducing number and the slightly higher blue solid line applying if the factor was instead *twice*.

Larger families in Ireland than in England have been *qualitatively* ascribed, without recourse to supporting modelling, to an early start to patrilineal surnames and by invoking an Irish custom of polygyny and by also pointing to the early hegemony of an important family. The medieval Welsh also practised early polygyny and it is not impossible that a persisting Celtic custom spilled over into early surname formation in parts of England. Celtic language and customs are often claimed to have existed in western England until they were driven back by the Normans; it is not inconceivable that they may have survived in pockets in England until times within the timescale of a widespread English custom for surname formation.

Evidence in the "early modern" historical period for the population of each particular surname is generally incomplete and accordingly unreliable. In so far as such evidence exists for Plant, however, the Hearth Tax data suggest that there may have been few more than about 50 Plant households around 1671, implying a significant growth in population size in the subsequent two centuries up to the 1881 UK Census. Though this evidence is of limited reliability, it points to the possibility of late enhancement in the growth of the Plant population corresponding to the blue lines in Figures 17 and 18. Conditions allowing increasing longevity between 1671 and 1881, might have had a small effect, by increasing the extent to which the total Plant population exceeded the number of reproducing males; this points to the green line in Figure 18, but even so the overall Industrial Age growth, irrespective of the demographic age profile, would apparently be an important factor.

To summarise, the contributory factors to prolific family growth may be varied. The fact that the benefits of the Industrial Revolution might have come to some regions earlier than others might be an important factor in the existence of some populous "single-origin surname contenders". This might go some way to explaining a contention of large single-surname families in West Yorkshire and Lancashire. However, large Industrial Age growth does not explain, in itself, all the differences that appear to arise in the sizes of surname families throughout England (Figure 6) even though it could well be part of the explanation. Early enhancement factors have also been considered, such as an existing fourteenth-century population due to an early start to a hereditary surname, or resistance to plague, or the fecundity of early polygyny, or favourable economic conditions in an early homeland. A particular finding is that chance, as well as favourable factors, is important in explaining the large size of a few families.

[45] There is a persistent view, dating back to a 1929 monograph by H J Hewitt, that Cheshire was subjected to excessive financial exploitation by the Black Prince but Hewitt's views are being challenged: English Historical Review (2007) CXXII (496) Appendix 12, p.89. Dr A M Tonkinson in *Macclesfield in the later fourteenth century* (Manchester, 1999) notes that peasant land holdings here were moderately substantial and that "Macclesfield Forest was an area associated with freeholding and customary tenure, with few people holding villein land; in fact, land could be freely demised or sold only on the payment of a customary relief of two year's rent".

Appendix. The Simulation Model

In this appendix we describe the simulation model used to explore the probabilities of outcomes associated with unusual population growth. The model keeps track of reproducing males. It 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[46] from the rate of population change in England in each generation. Population data were taken from Hatcher and Bailey[47] for the period from 1311 to 1541, and from Wrigley and Schofield[48] for the subsequent period.

Our model computes the number of male descendants of each progenitor by generating a random variable each generation representing the number of male offspring from each descendant in the current generation. This describes a type of branching process called a Galton-Watson process. An important property of such models is that the individuals do not interact with each other. That is, each individual in each generation procreates in isolation from the others; there is no competition for resources. Again, although this may not reflect the conditions of the real world, it does produce a sufficiently accurate simulation, because the effects of the competition are reflected in the overall population growth rate values.

Our objective, as described in the main body of the paper, is to simulate the behaviour of a population in which a particular surname has certain unusual properties, such as early polygyny, enhanced growth during some period, and so forth. We model polygyny by having each male in a generation in which polygyny is practiced 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 population of England in 1311 is conservatively estimated to have been approximately two to three million, implying a reproducing male population of approximately five hundred thousand. The simulation accordingly includes this number of individual simulated progenitors. Because these do not interact, their behavior can be calculated in any combination of population size and simulation runs. That is, for each parameter value, representing some exceptional property, we carry out a single simulation run in which all 500,000 surname lines have this value, and the effect is the same as if we had carried out 500,000 simulations in which one of them has this property.

[46] M A Pinsky and S Karlin, *An Introduction to Stochastic Modeling* (Academic Press, Boston, 2011).
[47] J Hatcher and M Bailey, *Modelling the Middle Ages* (Oxford University Press, 2001).
[48] E A Wrigley and R S Schofield, *The Population History of England 1541-1871: A Reconstruction* (Cambridge University Press, 1981).

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. We use a generation time of 30 years. This is 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 took the value of 30 as a round number that reflects a possibly shorter generation time during earlier centuries.