

By *patrilineage*, I mean the male line descendants of the earliest male ancestor, the patriarch, who lived within *genealogical time*—the time period within which genealogical research is possible and practical—roughly coincident with the time since written records began to be kept identifying individuals by name, and especially by hereditary surname.

Classifying Members of a Patrilineage on the Basis of Genetic Distance

FTDNA [rules out](#) a genealogical patrilineage relationship between two haplotypes with a genetic distance greater than 5 across the FTDNA 37-marker panel. Since I am prepared to allow for the inclusion in the same patrilineage of two people who are separated by a Genetic Distance (GD) of up to 10 or more, I have a lot of explaining to do.

In the first place, and most importantly, the FTDNA remarks on this subject are based on a conflation of the GD between the tested haplotypes of two living descendants, and the GD between a living descendant and the surname founder of his patrilineage. If GD represented that latter, then it's true that the accrual of as many as 6 mutations over 37 markers since the founder yields only a 50% probability that the tested descendant fits into a patrilineage founded 20 generations back (say about 1270), with a 10% chance that the ancestor could go back another 10 generations. But if the GD represents 3 different mutations incurred by each of two descendants since the founder, as is far more likely, then the 50% probability indicator drops to 10 generations (with the founder born about 1600), and only the very end of the long tail of the mutation frequency distribution points back as far as the 13th century. Under the circumstances, GDs of 6-8 might well be considered the norm for a patrilineage, not the exception, and in fact it is my experience that they are.

The FTDNA text more than hints at an understanding of this distinction, or at least it provides the basic rule for deciding when a haplotype with high GDs nonetheless fits into the patrilineage: when it has at least one lower GD relationship with any of the already accepted members. But a GD of 6-8 can stand on its own merit, and even support the inclusion of more remotely related patrilineage cousins, given certain other circumstances which no merely mathematical criteria can define.

Other Factors Bearing on Patrilineage Classification

The first of these is the simple fact that the candidate member has the same surname as the others. The TMRCA and GD calculators can't reasonably take this into consideration because many members of the patrilineage are likely to have other surnames, due to [NPEs](#) or for other reasons.^[1]

A recently published study by Turi K&J^[2] finds that for yDNA testing of bearers of 40 British surnames, selected randomly across England, an average of 41% of each set of surname bearers fell into a single most common patrilineal cluster for that surname.

Unfortunately, it is difficult to extrapolate this study directly to the largely American context in which the ancestries for most FTDNA surname projects are rooted, because most of the English surnames sampled are at least uncommon, if not rare, in America. I have accordingly analyzed the clustering percentages for three of the most common American surnames, and found that even for these clustering the percentages falling into the single most common cluster run in the high single

¹ Another reason might be an "Irish modal" descent from a prolific polygynist, like "Niall of the Seven Hostages".

² K&J-2009: "[Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames](#)", Turi E. K&J and Mark A. Jobling (2009)

digits and rise rapidly into the 20%+ range^[3], with the majority of the remainder falling into their own, smaller clusters.

What this all means is that if two British or American men bear the same surname, unless it is one of the 100 most common surnames, there is already at least a 20% chance that they fall into the same patrilineage *even before considering their test results*. And although I haven't done the Bayesian mathematics to quantify the additive probability effect of a common surname on the likelihood of a shared patrilineage, I think it's safe to say that this factor buys us an extra mutation or two at the least.

Second, we need to have reason to believe, in the case of a particular surname, that its origin really does go back far enough to accommodate a possible remote TMRCA. K&J-2009 predicate their conception of patrilineage (though they don't use that word) on a period of surname origination going back 700 years, and this is broadly consistent with the books I have read on this subject, for example, Richard A. McKinley, *A History of British Surnames* (1900), which is cited in the paper. McKinley and others also point out, though, that peak periods of surname origin have varied widely across the British Isles.

Surnames began to be used, at least by a few of the Norman gentry, in the wake of the Norman Conquest in 1066, and they came gradually into use in England by both gentry and commoners over the ensuing centuries. By about 1350 (the period when the Black Death wiped out as much as half the English population and set many of the others onto the road) surname use was general but by no means universal, and it didn't attain that status until the 1500s, perhaps given a bit of emphasis by K&J Henry the 8th's novel requirement that all baptisms henceforth be registered by local parishes.

Surname adoption has an even wider ranging history in Scotland, with some clan names being very ancient indeed, even pre-dating the Norman Conquest, while at the other end, many Scots didn't finally settle on their surnames until the mid-18th century! Surname adoption in Wales was similarly tardy, as anyone knows who has had occasion to spend time with the many Welsh Quakers who predominate in the records of the western Pennsylvania county of Chester before 1730.

The history of Irish surnames resembles that of their ethnic and genetic cousins, the Scots, only more so. There is strong evidence that a single individual Irish chieftain, by tradition Niall of the Seven Hostages himself fathered an enormous number of children many centuries before surnames began to be adopted. Thus Irishmen of many different surnames, some of them with exceedingly deep roots, test out as at least distant patrilineage cousins. On the other hand, surnames came into general use in Ireland much later than in England.

Given this extreme variability in the time of surname adoption, it seems to me that before we can seriously consider the possibility of a particular patrilineage being deeply rooted, and thus supportive of high GD numbers, we need to formulate some plausible theory as to when the particular surname first came into use, or at least of when it rose sufficiently out of obscurity to make an imprint on the records. Seeing the need for this, and actually doing something credible about it are two different things, but there are many books and encyclopedias which provide general guidance to surname etymologies, and beyond that, and more authoritatively, we have our genealogical indications.

And in fact this segues into the next consideration: what we know about the putative descendants from our genealogical research. For example, I have researched the early Robbs of Pennsylvania exhaustively in the records and determined that there were four Robb families planted here, within a forty mile radius well before the American Revolution, although they came to Pennsylvania over roughly a forty year period. Still this is small enough slice of space-time to create at least a strong

³ The surnames I analyzed were Walker (the 20th most common), Phillips (the 37th), and Perkins (the 184th).

suspicion that they were related before doing any yDNA testing. As it happens, we now have (to a high degree of confidence) yDNA samples from descendants of these four Robb families and it turns out that one pair of these haplotypes is related only at GD 8, which as we have already seen is well within the range of acceptability on its own merits. Still, if that number had been 10 instead, and given the common surname, would there have been much doubt that they still belonged to the same patrilineage?

Finally, I think it likely that most patrilineages go back at least two or three generations before their first hereditary use, which would buy us enough time, perhaps for another half mutation. I think that, first, because most moderns, even those most transient and rootless, know who their grandparents are or were, and know also of many of their first and even second cousins. In Britain, moreover, since “time immemorial” most people have had close ties to local places extending over many generations, and they have probably had at least a four-generation window of awareness about their ancestors.

In at least one area of late surname adoption, Wales, we know there was a naming system for keeping track of many previous generations, and also for identifying distant cousins. Indeed, the Welsh identified men more or less uniquely primarily by their descent, e.g. John ap Thomas ap Owen ap Thomas, and there is reason to believe that there were elders in each community who made a hobby of mapping these forests of “ap”s—pre-surname genealogists. And in Wales there was an additional, legal, reason for keeping track of remote cousins: under Welsh law, the land of a man who died intestate was parceled out in small pieces to very distant male cousins “of the blood”.

Thus, when surnames did come in, in most cases, it is likely that parents, aunts, and uncles (perhaps even grandparents), and more importantly, first, second, and maybe even third cousins adopted the same surname as the man who first passed it on to his children, thus pushing the patrilineage back a few more generations. I don’t suppose that this was always the case, but I think it at least represents a “fudge factor” in favor of deeper lineages, and higher GDs.