

TMRCA (Time to the Most Recent Common Ancestor, and GD (Genetic Distance) are both, in this context, measures of relatedness between two haplotypes (sets of tested ySTR marker values) drawn from genetic tests of the male yChromosome.^[1]

Mutations and Mutation Rates

The key constant in the TMRCA calculation, and in the interpretation of GD, is the average mutation rate of the tested markers, or better (for finer temporal resolution): the particular mutation rate of each of the markers that constitute the test panel. Unfortunately, though, the mutation rates of ySTR markers are known only very approximately, and in fact, the scientists know very little about the mutation process that affects them. ySTR mutations are conventionally assumed to be random, but as I have noted elsewhere, I think “random” is just a word meaning either “I don’t know the causes”, or, “the causes are too complex for me to work out and quantify at present”.

ySTR markers consist of a variable number of “repeats” of the same coding sequence of the “base pairs” A/C or G/T. The letters stand for the four nucleotides used to code the genome. Each letter of the genetic code, called a SNP, may be coded one of two ways: either as A or C, or as G or T. SNP mutations, which are much rarer than STR mutations represent a simple flipping to the alternate letter. No one knows what causes SNP mutations either, though there is no shortage of theories—or probably causes.

STR mutations, on the other hand (including STRs on the yChromosome) mutate by changing the number of repeats that make them up, usually by gaining or losing one repeat, in roughly equal proportions. Perhaps the currently favored theory of how ySTR mutations occur is by slippage in the meiotic replication process. One of the major items of controversy is whether ySTR mutations can increase or decrease by two or more “repeats” at a time, or at least whether this is a rare event, and if so, how rare. Other unresolved questions are whether ySTRs consisting of large numbers of repeats are more likely to mutate than shorter markers, and what the ratio of “up” to “down” mutations (whether a repeat is gained or lost) might be. It’s also far from clear (at least to me) that all these ySTR markers that we test really mutate independently of each other; this especially seems like an open question with respect to the “multicopy” markers about which I shall have more to say later.

One thing that seems crystal clear is that the mutation process is anything but random in any sense other than “indeterminate at present”. How can it be random when markers mutate at a different rate, some hundreds of times faster than others?

Under the circumstances, the current mutation rate estimates, which directly determine TMRCA estimates, must be considered provisional, but I hope and expect that over time the scientific and mathematical tools for interpreting haplotype variations across rapidly accumulating yDNA databases will become more precise, and thus render this data more useful to us genetic genealogists.

Genetic Distance and Mutation Models

Of the twin measures of relationship, Genetic Distance (GD) is the straightforward one. It means one of two closely related things: either it is simply the number of marker divergences between two haplotypes (assuming the “infinite alleles” mutation model) or it is the sum of the differences between the divergent marker values (assuming the “step-wise mutation model”). Over genealogical time (the span of anywhere from 300-800 years when documentary records might be found to support

¹ Haplotype (a contraction of “haploid genotype” may also refer to a broader sampling of the yChromosome to include SNPs (single event polymorphisms) that are indicative of the deep ancestral haplogroup. And there is an entirely different kind of haplotype: mitochondrial (extranuclear) DNA, inherited in haploid form from the mother.

conventional genealogical research), these models usually yield identical results because multistep mutations are so uncommon (a rough estimate might be once in 40-60 ySTR mutations).

The Problematics of TMRCA

If the calculation of GD is straightforward, estimating the TMRCA (Time back to the Most Recent Common Ancestor) of two haplotypes is another story. The calculation, using the formulas developed in the first few equations of Bruce Walsh's paper,^[2] depends on the number of markers and the number of mismatches (both known) and the mutation rates of those markers (for which we have only rough estimates). Where TMRCA is measured in years, rather than generations, there is also the indeterminate factor of the length of a generation.

The length of a generation is a straightforward empirical problem that can be dealt with briefly. It is often assumed, even in the scientific literature of population genetics, that 25 years is an adequate estimate for the length of a generation, but this is certainly very wide of the mark for the historic period with which genealogists are concerned. The best way to conceive of generational length as it relates to the patrilineages with which yDNA surname projects are concerned is the average age of the father when his average son is born. Various published studies, and an informal study that I have done on my own, all point to a value of around 34 years, which is the value I have adopted.^[3]

As for TMRCA, it is important to understand, and to remember, that while the calculation of TMRCA is precise, its accuracy depends crucially on the mutation rates assumed, and these are at present accurate only to about 15% plus or minus. And given the fact that the calculations produce a binomial distribution, which, unlike the classic bell curve, is quite heavily skewed out to the right (toward the distant past), even a small number of divergent mutations between two test subjects can produce significant probabilities that their actual common ancestor lived many hundreds of years earlier than the 50% probability peak that I use for the estimates shown in the TMRCA tables.

There is also, of course, a distribution towards the left, towards the present, so that there is an initially very small, but rapidly increasing chance that the ancestor lived closer to the present than the estimates show, but in the usual case, where at least enough research has been done to eliminate the possibility of a common ancestor back, say, 3-4 generations, a good bit of the cumulative probability for an earlier match must be lopped off the left (more recent) side of the curve. Thus, the numbers shown in the cells of the TMRCA tables displayed on the patrilineage pages of my DNA surname project websites, generally underestimate the actual time to the earliest common ancestor by at least a generation or two (the FTDNA Tip calculator, to be discussed anon, can correct for known genealogies).

It may seem counterintuitive that the mere additional knowledge that there could have been no common ancestor before, say 6 generations back, should change the underlying probabilities that they have a common ancestor back 8 generations, but the TMRCA estimates are derived by summing the separate probabilities of having a common ancestor 1, 2, 3... on up to 15-20 or more generations back, so even though the sum of the specific probabilities that the ancestor is exactly 1, exactly 2, ... and exactly 5 generations back, is a fairly low number, it does subtract from the cumulative probability of having a common ancestor 8 generations back.

Finally, there is a third, and perhaps most important consideration regarding TMRCA: the occurrence of the mutations upon which it is based are both very infrequent and subject to random

² Bruce Walsh "Estimating the Time to the Most Recent Common Ancestor for the Y chromosome or Mitochondrial DNA for a Pair of Individuals"; see SOURCES, at the end for citational specifics for this, and other sources.

³ Independently, Donn Devine, CG, had published an article on this for Ancestry magazine reaching the same conclusions. I recommend this article particularly for its citations to the literature; see SOURCES below for specifics.

chance. Thus, in the occasional case, perhaps one or two haplotypes out of ten, an anomalous number of mutations is going to occur, making it appear that the common ancestor between paired haplotypes goes much farther back into the past than is actually the case. For example, as many as 3, or even 4, mutations might occur where 1 is expected, causing the formulas to predict a common ancestor 400 or so years back in cases where the conventional research might seem to show pretty convincingly that the common ancestor lived just 200 years ago. Solid research generally trumps TMRCA estimates, but only if they are in the ballpark. If the number of mutations (the genetic distance) in this example were 6 or more, it would be time to carefully re-examine every link in the apparently proven ancestral chain to try to find the almost certain flaw. It's always possible, of course, that further scientific research will turn up special conditions where such anomalies as 6 mutations over a 200 year time span might reasonably occur, but given present scientific knowledge the odds of so many mutations occurring by chance are about 1 in 200.

The Assumptions and Estimates I Use in Calculating TMRCA

Calculation of a reasonably accurate TMRCA depends crucially on the mutation rates used, yet there is wide divergence in those reported rates. FTDNA, whose FTDNA TIP calculator appears to be the most advanced in some respects, presumably uses the following average mutation rate estimates (displayed in chart form at the 2004 ISOGG Convention):

"Estimating Mutation Rates of 37 Y-Chromosome Microsatellites Panel Mutation Rates

12	.00399
25	.00481
37	.00748"

It would appear that what FTDNA means by the table caption is that these "panel mutation rates" are the average per-marker per-generation mutation rates across each separate subset of the full 37-marker panel (where "generation" means "meiosis", or "transmission event"); that is, for example, the 25-marker panel rate of .00481 pertains not to the 25-marker panel as a whole, but just to markers 13-25. If this is so, then the average mutation rate per-marker per-generation for the 37-marker panel can be computed as

$$(.00399 * 12/37) + (.00481 * 13/37) + (.00748 * 12/37) =$$

.001294
 .00169
 .0024

 .005384 FTDNA Average Mutation Rate for 37-marker panel

while the average rate for the 25-marker panel would be

$$(.00399 * 12/25) + (.00481 * 13/37) =$$

.001915
 .002501

 0.004416 FTDNA Average Mutation Rate for 25-marker panel

These FTDNA rates, particularly the rate for the 12-marker panel run "hot" compared to those found in the published literature, and absent publication of the assumptions on which they are based, I consider their credibility in doubt. FTDNA is apparently clinging tightly to the "proprietary" algorithm coded into their FTDNATip calculator, and until they come clean about how it works, and

expose their claims to scientific scrutiny, I think the reported results of this calculator need to be regarded with suspicion.

I prefer instead to rely on the best *published* mutation rate estimates that have emerged so far, and those are the ones worked out by John Chandler.^[4]

The rates in CHANDLER-2006 are predicated on an 8340-haplotype sample drawn from the YSEARCH database, but he has also managed to incorporate, as a “calibration factor”, the cumulative rates found in GUSMAO-2005,^[5] which were derived largely from “gold standard” paternity test data. Paternity test mutation rate data are the most reliable because they virtually eliminate the uncertainties regarding ancestral derivation in comparing two haplotypes, by supplementing ySTR results with ySNP testing. GUSMAO for its part, published a set of mutation rates that were cumulative across most of the previously published mutation rate literature, however it must be said, first, that the GUSMAO data, even so, were limited, and only covered a rather small subset of the full 37-marker FTDNA panel used by CHANDLER.

Here are the CHANDLER rates for the three FTDNA test panels:

FTDNA 12-marker panel:	.00187	±	.00028
“ 25- “ ” :	.00278	±	.00042
“ 37- “ ” :	.00492	±	.00074

Note that the CHANDLER average mutation rates are cumulative for each marker panel, while the FTDNA rates shown in similar tabular form above apply just to the indicated subset of the full 37-marker panel, so that only the first, 12-marker, rates can be compared directly.

Even so, it is evident that there is a wide divergence from the FTDNA mutation rates, if only in the 12-marker panel rate, and in fact it turns out that most of the divergence lies there. CHANDLER’s data has shown certain of the 12-marker data to be even more “stodgy” than the GUSMAO data, especially marker DYS392. However, a couple of markers from the 13-25 marker set (454 and 455) also come in much slower than FTDNA reports. The 26-37 marker addition, however, which was designed to boost the average mutation rate, certainly lives up to its billing.

From private correspondence, I know that John Chandler would be the first to point out that, given the still relative scarcity of data, his estimates are very loose, exhibiting a very wide range of individual error, and even worse, the range of error rises for the faster (more likely to mutate) markers. Still, 37 (or actually 30 markers, treating the “multicopy” markers as single markers, as they are treated in all the published studies) are enough to command some respect for these overall averages.

Somewhat less well-grounded mutation rate estimates have been derived for FTDNA markers 38-67, and for the 11 additional markers used by DNAHeritage, Ancestry, and other testing companies. I’ve calculated mutation rate estimates for the most popular of these panels, and evaluated them comparatively, taking the costs of each test into consideration, [here](#).

The most important conclusion from this analysis is that the FTDNA 37-marker panel provides by far the “best bang for the buck” because the markers added when extending the original 25-marker panel to 37 run so “hot”. It’s incomprehensible to me that no other companies have followed this path since when cooking up new panels. For example, Ancestry just copied the old DNAHeritage panel when creating its flagship 43-marker test, instead of adopting the far better FTDNA 37-marker panel. And even FTDNA, in adding markers 38-67 to create a more sensitive, but also much more expensive test, would have served its customers better by adding just the most mutable markers from

⁴ CHANDLER-2006: “Estimating Per-Locus Mutation Rates”, John Chandler

⁵ GUSMAO-2005: “Mutation Rates at Y Chromosome Specific Microsatellites”, L. Gusmao, et al.

the 38-67 group, and from the additional 11 markers tested by SMGF, meanwhile dropping all the duds from its 1-37 panel. The result would have been a far more mutable panel at (probably) no additional cost over the present 37-marker panel..

Calculating TMRCA: YUtility and FTDNA's TIP Calculator

There are a number of free online TMRCA calculators of which by far the most sophisticated is [YUtility](#). I use this utility to produce my TMRCA charts, primarily because it allows one to calculate whole matrices of interrelationships for many haplotypes at once, and also for the extensive customization it supports. It's hands down the best utility for using the same data to produce Genetic Distance matrices. But YUtility (and all other free calculators) has one serious drawback compared with FTDNA's proprietorial Tip calculator: unlike Tip, it is unable to take genealogical knowledge into account. If one knows that the bearers of two tested haplotypes have no common ancestor back at least 3, or 6 generations, this has a material effect on the probability that there is an ancestor at 3, 6, or 9 generations. If one knows that there is no common ancestor for at least 6 generations, obviously the probability of a TMRCA at 3 or 6 generations is 0, while if one knows that there *is* a common ancestor back 6 generations, then the probability at 3 is 0%, and at 6 100%. Yet the YUtility charts ignore such knowledge and may estimate for all cases that there is a 50% probability of a common ancestor at 5 generations.

FTDNA Tip allows such knowledge to be factored in, but it has an equally serious disadvantage: it is only able to calculate TMRCA for a single pair of haplotypes at a time, and besides that, it can only be used to calculate estimates for matching pairs found through one's personal FTDNA web page.^[6]

Under the circumstances, I recommend using both, while reminding oneself not to take either set of estimates too seriously, because either can easily be off by 100-200 years or more.

Here are a few statistics to provide a feel for the mutation rate probabilities of the FTDNA 37-marker panel:

(1) The probability that exactly one or more of these 37 markers would mutate in any given transmission event (generation) is about 18.1% using the FTDNA average mutation rate, or 16.7% using the Chandler average mutation rate.

(2) Meanwhile, using Chandler for the individual marker mutation rates (FTDNA provides none), the probability that the fastest marker (either CDYa or b) would mutate is .03531, while the probability that the slowest (DYS426) would mutate is .00009: the fastest is 392 times more likely to mutate than the slowest! Because a few markers like CDYa&b are so fast, the marker closest to the mean (DYS439, with a mutation rate of .00477) is actually one of the fast markers. The ratios of these two fast markers (CDYa and 439) to the slowest (426) are therefore 392 and 53, respectively.

⁶ The Tip calculator is also able to take the individual mutation marker mutation rates into account, but the catch is that you have to trust FTDNA's undisclosed, never-published, mutation rate estimates, which, as a matter of fact, are suspiciously hot compared to those estimated from empirical data in the published studies.

Some Technical Complications with the Interpretation of Marker Values

The Various Testing Labs Fail to Embrace Standards, Which Fuzzes Several Issues

Regrettably, the half dozen or so ySTR testing companies have been slow to settle on a common set of testing standards, although some progress has recently been made in one area.

Different labs have been using different baseline numbers for reporting marker values, which has required some numeric conversions in order to compare results obtained at different labs. Last year, at the behest of a number of the leaders of the ISOGG community, but especially Katherine Hope Borges, several of the principal labs, including FTDNA's have agreed to use the NIST standards henceforth and to convert all their existing marker values accordingly. Fortunately this affects only a few markers (in the case of FTDNA at least), and just 2 in the 37-marker panel (DYS442, and YGATA-H4). Ancestry is apparently also on board this standards adoption, but it appears to be still in transition, so before making such comparisons it would be well to determine first whether the conversions have already been made, and to be aware that marker values downloaded earlier may have changed. SMGF (the Sorenson Molecular Genealogy Foundation) maintains a helpful [Marker Standards page](#) showing all the known lab standards, though how current this is, I do not know.

DYS389I & II

Another issue concerns the markers **DYS389I** and **DYS389II**. These appear in most testing panels yet there is a deep ambiguity in their nomenclature and meaning. While they are usually reported as independent values (FTDNA in particular is guilty of this), they are anything but. **DYS389II** represents an independent mutable allele, but it also incorporates the value of **DYS389I**, thus double-counting it. Even though FTDNA correctly subtracts the value of I back out of II in calculating TMRCA with its Tip calculator (Y-Utility does this too), misreporting it guarantees that most people who don't do this stuff for a living, but who like to wade into the nitty gritty and understand things for themselves, are going to innocently miscount the number of mutations across the marker panels where **DYS389I** has picked one up, and consequently are going to overcount genetic distance. For example, my own Robb haplotype has a divergent value at **DYS389I** from the others of my patrilineage, yet it matches on the underlying allele of **389II**, which nonetheless is shown as having a mutation as well. This underlying allele, by the way, has a name: **DYS389B**. But I had to poke around [at SMGF](#) to find out about this anomaly.

This ambiguity regarding **DYS389II** has been around for a long time. For example, in Table I of **KAYSER-2000**, showing mutations observed in that study, there are two reported mutations reported for **DYS389II**, one from allele value 16 to 17, and one from 17 to 18, and each value is annotated with the note: "Excludes **DYS389I**". Why FTDNA and others persist in reporting these results incorrectly, leading many, I am sure, to derive the wrong values for genetic distance, and incorrect estimates for TMRCA, I cannot understand.

Multicopy Markers

The final problem, which is potentially much more serious, is the problem of the so-called "multicopy" markers. These are a special class of ySTRs that are not only composed of multiple repeats of particular nucleotide strings, the STRs themselves exist as multiple copies, mostly two such copies, but in one case there are four or more. The markers I am speaking here of are the ones with suffixes, "a", "b", "c"..., and the four-copy marker, **DYS464a-d**, occasionally extends to additional copies—**DYS464d**, e.... The problem is that while separate independently mutating marker values are reported for each of the suffixed components of these multicopy markers, these values are in a certain sense both interdependent and indeterminate.

The reason for this is that the lab tests of these markers produce values for each of the two (or four or more) suffixed subsites, without being identifying the order in which they occur.⁷ If the composite values for DYS464a-d were reported as 15-15-15-15 in one member of a patrilineage, and 14-15-15-15 in another, this does not necessarily mean that it was the microsatellite designated by DYS464a that mutated: it might just as well have been DYS464b, c, or d. It is merely conventional to list these four related values in ascending order. And because we don't know the order in which the components occur, there is no legitimate way to compare them one-to-one between haplotypes.

It appears that different labs and calculation utilities handle this anomaly in different ways. I have learned that both FTDNA Tip and YUtility treat the multicopy markers DYS464 and YCAAI as single markers with summed mutation rates, according to the "infinite alleles" mutation model, which treats any number of mutations to a marker as a single mutation. For reasons that I am still trying to determine, the other multicopy markers (DYS459a-b and CDY a-b in the 37-marker panel, and DYS395S1a-b and 413a-b in the 38-67 marker set) are treated according to the "stepwise mutation model" with each step of difference treated as a separate mutation. In general, the scientific studies support the stepwise approach, for example GUSMAO-2005, which consolidated data from many other studies found only one multistep mutation out of 54, and it seems to me that this adjustment is best regarded as a testing artifact, although that multistep mutations do occasionally occur is undeniable.

DYS464 is a particularly grievous problem, because it (collectively) constitutes 4 markers of the 37-marker panel, and even taken one component at a time is one of the fastest-mutating markers and therefore potentially most useful for genetic genealogical purposes. By ignoring the possibility that it might have experienced more than one mutation during genealogical time, which, given its fast mutation rate is bound to happen occasionally, it introduces an unfortunate element of uncertainty into GD and TMRCA.

For example, consider the two sets of values 14-15-15-16 and 14-15-16-16. Let us suppose that the former is the original genotype that has somehow mutated to the latter set of values, and let us assume further that the order in which the original was reported corresponds to the actual order of the microsatellites. The latter set of values might have arisen by an "up"-mutation at either position b or c of the original, or from a two-step mutation at position a, and unless we are quite sure that only one mutation could have occurred at this volatile multicopy marker during the timespan back to the original genotype, there are many other far more complex permutations. In fact, given the current controversy over whether two-step mutations are other than rare, there is no way to tell just how many mutations may have occurred at this 464. What's more, the very existence of these multicopy markers suggests that there are genetic mutational mechanisms at work of which we know nothing at present.

In fact it seems entirely plausible to me that the ambiguities of these multicopy markers may have disguised an inherent interdependence, which would invalidate the components as *independent* markers. In support of this conjecture, I note that while SMGF (which tabulates the combinations of allele values for 464), [is presently reporting](#) no fewer than 658 value sets (including sets for which there are more than 4 values: DYS464e,f, etc., and many fractional readings), most of them with microscopic probabilities but clustering on a double handful of scattered set values with small but significant probabilities, e.g. 15-15-17-17 constitutes about 10% of their database, while 15-16-17-17

⁷ There is one exception to this rule. For DYS385a&b there is a special test, called the "Kittler test" for determining the order of the components, a & b. However, this is not part of the standard testing regimen and most prospective testees aren't made aware that the Kittler test option exists.

is 2.7%. This looks like anything but a chance, bell curve like, distribution, which has evolved over time by more or less even dispersion. It appears that other unknown factors are working here.

The same theoretical objections apply to the other multicopy markers, but because these involve only two sites, and thus have a decreased likelihood of ambiguous multiple mutations, it is more reasonable to treat them as independent markers.

No wonder such authorities as Ken Nordvedt, a nuclear physicist whose hobby is population genetics, and who is applying sophisticated mathematical techniques toward working out “modal haplotypes” for deep lineage haplogroup subclades, leaves these multicopy markers out of his calculations altogether. So does SMGF, for that matter.

There is no doubt in my mind that FTDNA’s way of handling DYS464 is the best under the circumstances. As genetic genealogists, focused on the “short term” (the last 800 years or so) we can’t afford to dispense with these important markers.

Markers Can Not only be Ambiguous, They Can also be too Fast, or too Slow

CDY mutates so readily that it, too, can distort the GD and TMRCA calculations. If we consider that typical surname patrilineages go back no more than 20 generations, the odds that one of the CDYs would mutate in a particular haplotype over that period would be about 50%, and if this should happen early in the descent, there could be almost a 25% chance that it will mutate back to its original state during the remaining generations, thus wiping out two mutations that should have been counted. I have seen at least one probably instance of this in data I have looked at, and a rough estimate of the chance of this happening in a shallower lineage of say 14 generations might be around 5%, or one haplotype in 20

The ultraslow markers, on the other hand,^[8] are so unlikely to mutate over 20 generations (there is only about 1 chance in 500 for each, or 1 in 50 that any will mutate) that when they occasionally do the event is best dismissed as an anomaly.

We are using marker mutation rates as a clock to measure the passage of genetic time, and the periods for the CDYs and the ultraslow markers are simply inappropriate for the purpose, although the ultraslows are still useful for differentiating patrilineages, and the CDYs for differentiating sublineages.

That leaves 28 appropriately calibrated markers out of the 37, and most of the mutations that occur affect the nine most mutable of these: DYS576 ($p=.0102$), 449 ($p=.0084$), 458 ($p=.0081$), 570 ($p=.0079$), 456 ($p=.0074$), 464a-d ($p=.006$), 439 ($p=.0048$), 607 ($p=.0041$), 460 ($p=.0040$).

DYS576, the fastest of these markers, has about a 20% chance of mutating over 20 generations, while if we consider $p=.007$ to be a typical mutation rate for these fast markers, the chances of any particular one mutating are about 12% over 20 generations, and most of the time the MRCA of the haplotype set we are considering probably doesn’t go back that far.

A Final Word

Our existing knowledge of the mutation process is still quite primitive and it is to be hoped that with better scientific data, lower testing prices, and better tests that we will get a better fix on mutation rates and on the vagaries of the process. If FTDNA were to construct the sort of test I’ve recommended above (dropping the inappropriately calibrated dud markers from the 37-marker panel, and replacing them with fast markers from the remaining testable set, I believe that the accuracy and usefulness of this test for genealogical purposes would be greatly improved.

⁸ DYS388, 392, 393, 426, 438, 454, 455 in the FTDNA 37-marker panel, and there are many others in the 38-67 group.

SOURCES:

CHANDLER-2006

“Estimating Per-Locus Mutation Rates”

John F. Chandler

(*J Genet Geneal*, 2:27-33)

<http://www.jogg.info/22/Chandler.pdf>

DEVINE-2005

“How Long is a Generation”

Donn Devine, CG

Ancestry Magazine (Sep2005)

<http://www.ancestrymagazine.com/2005/09/research/how-long-is-a-generation/>

GUSMAO-2005

“Mutation Rates at Y Chromosome Specific Microsatellites”

L. Gusmao, et al.

(*Human Mutation* 26(6):520-528, 2005)

http://www.gep-isfg.org/documentos/2005_GEP_HumMut.pdf

KAYSER-2000

“Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs”

Manfred Kayser, et al.

(*Am. J. Hum. Genet.* 66:1580-1588, May2000)

http://www.familyreedna.com/pdf/Kay_AJHG_2000.pdf

KING-2009

“Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames”

Turi E. King and Mark A. Jobling

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Email: maj4@leicester.ac.uk

Pre-Publication Copy

<http://mbe.oxfordjournals.org/cgi/content/abstract/msp022>

McKINLEY-1990

Richard A. McKinley

A History of British Surnames (EssexENG: Longman Group Limited, 1990)

WALSH-2001

“Estimating the Time to the Most Recent Common Ancestor for the Y chromosome or Mitochondrial DNA for a Pair of Individuals”

Bruce Walsh, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

(*Genetics* 158: 897-912, Jun2001)

<http://www.genetics.org/cgi/reprint/158/2/897.pdf>