Journal: www.jogg.info
Originally Published: Volume 4, Number 2 (Fall 2008)
Reference Number: 42.005

MORE REALISTIC TMRCA CALCULATIONS

Author(s): Kenneth Nordtvedt

Journal of Genetic Genealogy, 4(2):96-103, 2008

More Realistic TMRCA Calculations

Kenneth Nordtvedt

Abstract

The traditional estimates of age back to the most recent common ancestor (MRCA) for a pair of present-day Y-STR haplotypes implicitly assume that the haplotype possibilities for that MRCA have the same chance of being found in any past generation's population, i.e., that population's haplotype frequencies are constant through time. In reality, haplotypes are found to cluster near clade modal haplotype with high frequency, and this requires those frequencies to be changing in time as the clusters tend to widen and their central peak frequencies fall. Using the full Bayesian rule of probabilistic inference, these time-changing haplotype frequencies are taken into account, and as a result major changes to the age estimates for MRCAs are derived. In one extreme, if one of the favored MRCA haplotype choices is the modal (founding) haplotype of the clade to which the pair of present haplotypes belong, the time estimate back to the MRCA is doubled from the traditional estimate. If, on the other hand, the pair of haplotypes have all their favored choices for MRCA haplotype many steps of mutation from the clade modal haplotype, the MRCA time estimate can be pushed toward the present and become younger than traditional estimates. This modified method for estimating a TMRCA can be viewed as triangulation in which a clade's founding haplotype joins with two present-day haplotypes in the process of best-estimating the location of an intermediate node (the MRCA being sought) in a portion of the clade's phylogenetic tree.

Introduction

Many genetic genealogists eventually employ a time-to-most-recent-common-ancestor (TMRCA) tool to estimate how far back in time the common ancestor existed for two Y-STR haplotypes obtained from extended family members. Y-haplotypes consist of some number N of single tandem repeat (STR) segments of y-chromosome DNA material (henceforth called simply markers) whose number of repeats of specific DNA building blocks can be measured. The number of repeats for each STR are almost always faithfully inherited by each son from his father, but an STR occasionally mutates at a father-to-son transmission with its own marker mutation probability ranging from 1/100 down to 1/10,000 per transmission for the commonly measured STRs. Two present-day haplotypes which descend from a common ancestor who lived G generations ago will therefore tend to develop differences in some of their marker repeat values, and the number increases with the size of G—the generational time back to their most recent common ancestor (MRCA). The number of marker mutational differences, n, accumulated between two haplotypes will become on average

$$n = 2MG$$

with M being the sum of marker mutation rates.

Address for correspondence: knordtvedt@bresnan.net

Received: August 7, 2008; accepted: September 18, 2008.

Therefore, by observing n, the genetic distance, between the pair of haplotypes, and having a knowledge of M, G can be inferred; or, in the absence of knowledge of the total mutation rate, M, ratios of TMRCAs for different haplotype pairs can be inferred from ratios of the observed n values. The probabilistic nature of STR mutations, however, rule out a determination of a very specific TMRCA; in reality there will sometimes be more and sometimes less than the average number of mutations occurring between the haplotypes since their MRCA, so a probability distribution for the number of generations to the TMRCA is what a particular observed n allows us to infer. In carrying out the more complete analysis for the traditional TMRCA probability distribution, we will bring to light an unnecessary, and indeed unrealistic, implicit assumption about the ancestral distribution of haplotypes which goes into the standard analysis, and then in the next section move on to a more realistic consideration of the TMRCA estimation problem which results in major changes to the inferred TMRCA probability distribution.

The standard mutation model for the N markers here employed is as follows: each marker i, having its own mutation rate $m(i)$, is transmitted from father to son unchanged in its repeat number with probability $1-m(i)$, or is increased by one repeat unit with probability $m(i)/2$, or is decreased by one repeat unit with probability $m(i)/2$, with these rules independent of marker repeat number. STR mutation behavior is probably more complex than the standard model and could be incorporated into TMRCA analysis, but such model refinements have not yet been measured and quantified to any degree by the research community.

Review of the Traditional TMRCA Solution

The basic TMRCA problem for two present-day N-marker Y-STR haplotypes $Hap(y)$ and $Hap(Y)$ is to determine the probability distribution for the number of generations, G, back to their most recent common ancestor. Such a distribution will show what is the most likely TMRCA for the haplotype pair, and also show how broadly that probability spreads above and below the most likely TMRCA, i.e., the confidence interval for a TMRCA estimation. In addition, the TMRCA analysis also identifies what the MRCA's haplotype, $Hap(k)$, is most likely to have been. A specific case is first discussed in this paper, and then we work toward the general situation.

Suppose the two present-day haplotypes, $Hap(y)$ and $Hap(Y)$, differ by one repeat unit on each of four markers and have identical alleles at their other N – 4 markers.[1] Then with a, b, c, d being the alleles for $Hap(y)$ at the four markers where the two haplotypes have one-step allele differences, and A, B, C, D being the corresponding alleles for $Hap(Y)$, there are sixteen first-tier choices for the common ancestor haplotype $Hap(k)$. First-tier haplotypes for the MRCA are those which can reach the two present haplotypes $Hap(y)$ and $Hap(Y)$ through the minimum number of mutations—four in this case being discussed. Their probabilities of producing $Hap(y)$ and $Hap(Y)$ in later generations are substantially higher than other choices for the MRCA haplotypes, so consideration is restricted to them.[2]

$$\begin{aligned}
&\boldsymbol{Hap(k1) = \{(a,b,c,d)\}}\\
&\boldsymbol{Hap(k2 - k5) = \{(A,b,c,d),\ (a,B,c,d),\ (a,b,C,d),\ (a,b,c,D)\}}\\
&\boldsymbol{Hap(k6 - k11) = \{(A,B,c,d),\ (A,b,C,d),\ (A,b,c,D),\ A(a,B,C,d),\ A(a,B,c,D),\ A(a,b,C,D)\}}\\
&\boldsymbol{Hap(k12 - k15) = \{(a,B,C,D),\ (A,b,C,D),\ (A,B,c,D),\ (A,B,C,d)\}}\\
&\boldsymbol{Hap(k16) = \{(A,B,C,D)\}}
\end{aligned}$$

These sixteen first-tier MRCA haplotype alternatives consist of: one identical to either $Hap(y)$ or to $Hap(Y)$; then four choices for the common ancestor's haplotype that are one step of mutation from $Hap(y)$, and also four choices for being one step from $Hap(Y)$; the remaining six choices for the MRCA haplotype will be two steps from both $Hap(y)$ and $Hap(Y)$. Under feigned or real maximal ignorance of any other information about the problem, all sixteen of these choices have equal probability of being the MRCA's haplotype.[3] The probability that the haplotypes $Hap(y)$ and $Hap(Y)$ descend from any one of those sixteen first-tier MRCA haplotypes from G generations in the past then calculates to be:

$$Prob[Hap(y),\ Hap(Y) \mid Hap(k), G] = \frac{m(1)m(2)m(3)m(4)}{16}\ G^4 e^{-2MG} \qquad (1)$$

Each of the four required and specific markers mutate once, either up or down, at their individual rates $m(1)/2$, . . . , $m(4)/2$, regardless of which of the sixteen haplotypes was the actual MRCA; and each of the four mutations had G generations (chances) for happening. The final exponential factor in Equation (1) expresses the probability that the N markers of the haplotypes with total mutation rate M,

$$M = \sum_{i=1}^{N} m(i), \qquad (2)$$

otherwise did not mutate over the $2G$ generations of branch length connecting $Hap(y)$ to $Hap(Y)$ via $Hap(k)$ (see Figure 1). The resulting probability is seen to vary with number of generations into the past as $G^4 e^{-2MG}$, reaching a peak at the location $G=2/M$, thereby defining the most likely G, but the probabilities of different possible G outcomes spread substantially above and below the most likely G, as shown by the distribution curve plot in Figure 1.

1 Allele differences of more than one repeat could be considered here, but the added complexity of the discussion without much added to the essential conclusions does not justify doing so in this introductory paper.

2 Each additional mutation in a tree connecting a MRCA haplotype to the two present-day haplotypes costs a factor of $m(i)G$ in probability, with m(i) being the mutation rate of the additional marker and G being the number of generations back to the MRCA. MRCA haplotypes other than the first-tier ones would need at least two additional mutations to occur, and the factor $m(i)G$ is much less than one for recommended applications.

3 In this decade's early years when genetic genealogy was in its infancy, perhaps it made sense to promote the simplified traditional TMRCA model, which neglected information about the variation in the frequency of haplotypes in the ancestral populations from which MRCAs must be chosen.

Journal of Genetic Genealogy, 4(2):96-103, 2008

Generalizing to the case of n one-step allele differences between $Hap(y)$ and $Hap(Y)$, the probability of the pair being reached from any of the first-tier choices of MRCA haplotype, $Hap(k)$, becomes:

$$Prob[Hap(y), Hap(Y) \mid Hap(k), G] = \frac{m(1)m(2)\ldots m(n)}{2^n} G^n e^{-2MG} \qquad (3)$$

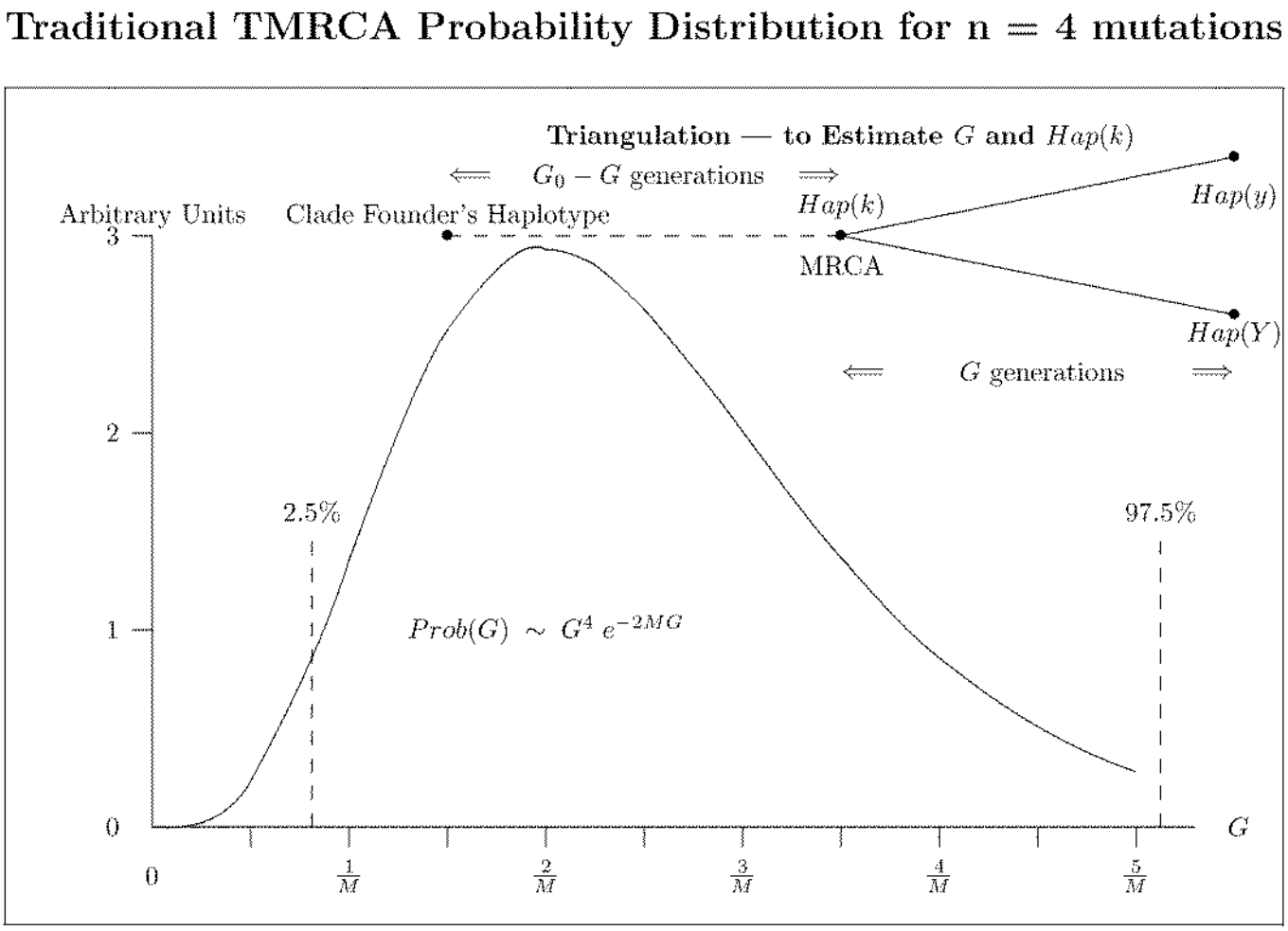

Figure 1 Under traditional analysis which neglects the variation in frequency for different haplotypes being present in the ancestral population, given two present-day haplotypes, $Hap(y)$ and $Hap(Y)$ with four markers showing mutational differences and M being the sum of marker mutation rates for all their N markers, the illustrated curve gives the relative probability that their MRCA occurred G generations ago, regardless of which of the sixteen first-tier MRCA haplotypes $Hap(k)$ is considered. The distribution peaks at the most likely $G = 2/M$, and 95 percent confidence interval boundaries are shown at $.81/M$ and $5.12/M$. The two solidly drawn branch lines from the present-day haplotypes and converging on the MRCA represent the traditional analysis. This paper derives the major changes to this picture that result due to consideration of the complete Bayesian rule which takes into account time-dependent variation in haplotype frequencies in the ancestral populations. The dashed branch line from MRCA back to the clade founder's haplotype reveals the triangulation which becomes part of this more realistic estimation of TMRCA. The probability distribution for the MRCA and his most likely haplotype is now determined by the three known and encompassing "facts"— $Hap(y)$, $Hap(Y)$, and the clade founder's haplotype.

With the probability distribution's peak (most likely estimate) occurring at

$$G = \frac{n}{2M} \qquad (4)$$

In comparing the chances of the TMRCA occurring at various numbers of generations ago, an implicit assumption leaked into the above discussion—that the chances of finding each of the specific first-tier haplotypes for the MRCA in the ancestral population, was independent of time in the past. A static and uniform history of haplotype frequencies in past populations is not only unrealistic, studies of Y-STR haplotype databases in the last few years by both the genetic genealogy community and others find that a population's haplotypes are typically clustered around the clade modal haplotype with that modal haplotype occurring at an unusually high frequency. At the beginning of the clade, the founder's haplotype was the only one present and had a frequency of 1.0. This change to a broader distribution of haplotypes over time, necessarily implies that the haplotype frequencies are changing in time--contrary to the implicit assumption. Consideration of this more complete picture requires us to step back and employ the full structure of Baye's Theorem of probabilistic inference, whose general form states:

$$Prob[A \mid B] = Prob[B \mid A]\ Prob[A] \ \text{(times a normalization factor)} \qquad (5)$$

This equation states in words: The probability of A, given facts B, equals the probability of B given A as facts, multiplied by the a-priori probability that A is true. And the right hand side of this Bayesian relationship is normalized, or when normalization is difficult or impossible, it can be used to produce ratios of probabilities, $Prob[A(i) \mid B] / Prob[A(j) \mid B]$. The new ingredient, previously neglected in the traditional TMRCA approach, is the *a priori* probability distribution $Prob(A)$. Applying the full Bayesian formulation to our TMRCA problem, we see that an additional factor of $Prob[Hap(k), G]$ must be included:

$$Prob[Hap(k), G \mid Hap(y), Hap(Y)] \sim Prob[Hap(y), Hap(Y) \mid Hap(k), G]\ Prob[Hap(k), G] \qquad (6)$$

The probability that a haplotype, $Hap(k)$, *is* that of the MRCA who lived G generations ago, given the two present-day descendant haplotypes, $Hap(y)$ and $Hap(Y)$, is proportional to the probability that the two haplotypes, $Hap(y)$ and $Hap(Y)$, will result from being descended from a MRCA with haplotype $Hap(k)$ living G generations ago, then multiplied by the probability there was a haplotype $Hap(k)$ in the population G generations ago.

Taking the Time-Varying Past Haplotype Population into Account

When seeking the TMRCA for $Hap(y)$ and $Hap(Y)$, the two haplotypes should be of the same clade. [4] If the two are not of the same clade, then it is probably more profitable to investigate the estimated ages for their different clades, as the MRCA for haplotypes from different clades will be from an era prior to the age of at least one of the clades from which the pair descend. Large clades often have ages of thousands of years, pushing the age for the MRCA far beyond a genealogical time frame.

It is important to identify the most recent clade from which the haplotype pair of interest descended. How the pair of haplotypes compare with their clade's modal haplotype over the full set of N markers plays a key role in the modification of TMRCA estimates which follow, so the clade's modal haplotype becomes an important ingredient in the following discussion.

The inclusion of a-priori information about the frequencies of various haplotypes being present in past populations, as required by the full Bayesian formulation of our problem, produces a modified expression proportional to the probability for the MRCA occurring G generations ago:

$$Prob[Hap(k), G \mid Hap(y), Hap(Y)] \sim G^n e^{-2MG}\ f[Hap(k, G)] \qquad (7)$$

4 A haplotype clade consists of haplotypes which descend from a discernable common ancestor. In the absence of an SNP (single nucleotide polymorphism) tag for the clade, it is identified by the clustering of the members's Y-STR haplotypes near a founding haplotype motif. Haplogroups are clades for which a SNP tag indicates the common ancestry of the haplotypes.

Equation (7) simply has the added factor of the frequency for finding the MRCA haplotype $Hap(k)$ G generations ago. Suppose two first tier alternatives for the MRCA's haplotype—$Hap(k)$ and $Hap(K)$—are under comparison for being the MRCA haplotype. The relative probabilities that one or the other is the haplotype of the MRCA is the relative size of their presence in the population of G generations ago, because each would have had an identical probability (given in Equation (3)) of producing today's $Hap(y)$ and $Hap(Y)$.

A particularly interesting situation is when one of the first tier alternatives for the MRCA haplotype for $Hap(y)$ and $Hap(Y)$ is the modal haplotype of the clade from which the pair descend. For reasonably young clades,[5] the frequency of the clade modal haplotype being present in its descendant population after $G_0 - G$ generations from its founding is given by Equation (8):

$$f[Hap(k{=}modal),\ G] \approx \prod_{i=1}^{N} [1 - m(i)(G_0 - G)] \ \approx\ e^{-M(G_0 - G)} \qquad (8)$$

As time approaches the present, this is a diminishing frequency for finding $Hap(k{=}modal)$, although the frequency of a clade's modal haplotype will generally remain the largest frequency in the clade cluster. This thereby makes the clade modal haplotype the most likely MRCA haplotype among the first-tier choices. Inserting this falling probability into Equation (7) gives the modified overall probability for a MRCA haplotype being present G generations ago and then producing today's $Hap(y)$ and $Hap(Y)$:

$$Prob[Hap(k{=}modal),\ G \mid Hap(y),\ Hap(Y)] \ \sim\ \ G^{n}e^{-MG} \qquad (9)$$

The diminishing frequency with time from the clade's founding for the clade's modal haplotype in the descendant population profoundly alters the estimate for the most likely value of G. Comparison with Equation (3) shows that the time scale parameter $1/(2M)$ has been doubled to $1/M$ because of the increasing chances of finding the modal haplotype further back in time. The effect on TMRCA estimation in this case is both to double the most likely TMRCA and also to double the high and low age boundaries for any confidence intervals one chooses to bracket around the most likely estimate. The traditional curve in **Figure 1**, for example, simply has its time axis scale doubled.

When the clade modal haplotype is one of the first-tier alternatives for MRCA haplotype, this haplotype choice for the MRCA will stand above the alternatives in liklihood, and the most likely TMRCA for this choice, along with its confidence interval boundary points, are doubled from what traditional TMRCA analysis yields.

In general, however, each haplotype $Hap(k)$ among the first-tier alternatives will have a non-zero total number of steps of difference, $D(k)$, from the clade modal haplotype. For instance, if some of the $N - n$ markers for which $Hap(y)$ and $Hap(Y)$ have identical allele values, nevertheless differ from the clade modal haplotype, that serves as a minimum floor for the MRCA haplotype distance from the modal haplotype. As one ranges over the 2^n first-tier alternatives for the n markers where $Hap(y)$ and $Hap(Y)$ differ, $D(k)$ can only remain at or increase from that floor. A bit more modeling is needed to find how large and how varying in time are the frequencies for those choices of MRCA haplotype where $D(k) \neq 0$. For young clades their frequencies will be proportional to:

$$f[Hap(D(k)),\ G] \ \sim\ \frac{m(1)m(2)\ldots m(D(k))}{2^{D(k)}}\,(G_0 - G)^{D(k)}\,e^{-M(G_0 - G)} \qquad (10)$$

where the factor $m(i)(G_0 - G)/2$ is the probability that the respective markers mutated from the MRCA's marker values over the time interval $(G_0 - G)$ generations, G_0 being the number of generations back to the clade founding. This frequency function, substituted into Equation (7), yields Equation (11), showing the G dependence of the probability curves for TMRCA when the choice of MRCA haplotype k is at a genetic distance $D(k)$ from the clade modal haplotype:

5 A "reasonably young" clade means one that is sufficiently young that multiple mutations of the same marker, leading to some back mutations to the modal value, is a rare occurrence.

$$Prob[Hap(k),\, G \mid Hap(y),\, Hap(Y)] \sim G^{n}(G_0 - G)^{D(k)}\, e^{-MG} \qquad (11)$$

The frequency distribution given by Equation (10) does two things to change the overall probability distribution for TMRCA of Equation (7); the exponential factor pushes the distribution to higher G values, while the factor $(G_0 - G)^{D(k)}$ pushes the distribution to lower G values. The peak of this resulting probability distribution yields the most likely TMRCA estimate and moves to:

$$G(k) = \frac{n}{M + \dfrac{D(k)}{[G_0 - G(k)]}} \qquad (12)$$

For sufficiently large genetic distances of the MRCA haplotype from the clade modal haplotype, we will have $D(k) > M[G_0 - G(k)]$, so the most likely TMRCA will be closer to the present than the traditional analysis result $n/(2M)$. The confidence intervals quoted in fractional terms remain the same or are narrowed. A good surrogate for the standard deviation of the probability distribution is given by the probability distribution divided by its second derivative (with respect to G) evaluated at the peak of the distribution. This yields Equation (13), the distribution's standard deviation estimate (in units of G):

$$\frac{\sigma}{G} \approx \frac{1}{\sqrt{n + D(k)G^2/(G_0 - G)^2}} \qquad (13)$$

Of course the actual distribution function given by Equation (11) can be plotted and confidence intervals determined for various choices of n, $D(k)$, G_0, and M. The basic $1/\sqrt{n}$ dependence of the distribution's standard deviation highlights the crudeness of the TMRCA tool for estimating when a recent common ancestor lived; TMRCA becomes more interesting for the deeper ancestral estimates with greater differences n between the haplotype pairs. An example application of the modified TMRCA estimate tool derived above is made in this paper's Appendix for a pair of haplotypes from Scotland.

If one wants to estimate TMRCA from a more empirical standpoint, avoiding the analytical estimates made above for the frequencies of clade haplotypes of various distances $D(k)$ from the clade modal (founding) haplotype, then actual haplotype frequencies found in appropriate databases can be used to make these estimates. The change per generation in the various haplotype frequencies can be expressed in terms of the frequencies themselves, with any particular haplotype's rate of change determined by its $2N$ nearest neighbor haplotype frequencies as well as its own frequency as shown in Equation (14) below. The left side of Equation (14) represents the change in frequency of haplotype $Hap(k)$, the first term on the right side represents the loss due to the haplotype in question mutating to any of its neighboring haplotypes, while the last term on the right represents the gain due to all neighbors $n(k)$ mutating to the haplotype.[8]

$$\frac{d}{dG} f[Hap(D(k), G] \sim -M f[Hap(k), G] + \sum_{n(k)=1}^{2N} m(n(k))\, f[Hap(n(k)), G]/2 \qquad (14)$$

Using this expression for rate of change of haplotype frequency in the population, and setting the derivative of Equation (7) to zero, the probability peak is found when

$$0 = \frac{n}{G(k)} - 2M + \frac{d\{\log f[Hap(k), G]\}}{dG} \qquad (15)$$

yielding the empirically-based estimate of most likely TMRCA as shown in Equation (16):

$$G(k) = \frac{n}{M + \displaystyle\sum_{n(k)=1}^{2N} \frac{m(n(k))}{2} \frac{f[Hap(n(k)), G]}{f[Hap(k), G]}} \qquad (16)$$

$m(n(k))$ is the mutation necessary to convert $Hap(k)$ into the neighboring haplotype $Hap(n(k))$. Note that if the frequency of the MRCA haplotype, $f[Hap(k)]$, is substantially greater than the frequencies of its $2N$ neighbors, as will be the case if it is the same as the clade's modal haplotype, then the above reproduces the result of Equation (9), which doubles the estimated age for that choice of MRCA haplotype.

8 Consequently, time-independent haplotype frequencies in a population require uniformity of frequency across the haplotypes.

A very large database of clade haplotypes will be necessary to obtain good frequency determinations for the extended whole-haplotype frequencies. If the clade shows no further sub-clade structure, a good approximation to the haplotype frequencies can be made from the clade's observed individual marker frequencies, which can be obtained from a smaller database. The product rule of composition for independently mutating markers can then be used to infer the extended whole-haplotype frequencies as shown in Equation (17).

$$f[Hap(k)] \approx \prod_{i=1}^{N} f[r(i,k)] \qquad (17)$$

with $f[r(i,k)]$ being the frequency of the *ith* marker having the repeat count $r(i,k)$ equal to that for the haplotype $Hap(k)$. Equation (15) then simplifies to Equation (18):

$$G(k) = \frac{n}{M + \sum_{i=1}^{N} m(i) \frac{f[r(i,\,k) + 1] + f[r(i,\,k) - 1]}{2 f[r(i,\,k)]}} \qquad (18)$$

Allele frequency distributions are readily calculated from good databases of Y-STR markers, but these frequency distributions necessarily represent the present. The present-day distributions can probably be used without modification for a moderate number of generations into the past. Or, they can be corrected from th the present into the past using the basic mutation model equations for each marker, as shown in Equation (19):

$$\frac{d}{dG} f[r(i),\, G] = \frac{m(i)}{2}\left(2f[r(i),\, G] - f[r(i) + 1,\, G] - f[r(i) - 1, G]\right) \qquad (19)$$

A Closing Note of Caution

A TMRCA estimate for a pair of present-day haplotypes is a fairly blunt tool at best, and it is easy to read too much into it. The statistical confidence intervals for such estimations are very wide, even if very high confidence in the underlying mutation model for the markers were at hand, which is not yet the case. But with that caveat, if the tool is to be used at all, it should not start from the very beginning with up to 100 percent error due to neglect of using information on the particular haplotypes involved. Such information pertinent to the haplotypes under examination is now readily available in the present Y-STR databases—databases that are growing rapidly, especially for certain regions of the world.

References

Nordtvedt K (2008) Founder haplotypes for Y-Haplogroup I, varieties and clades. URL: http://knordtvedt.home.bresnan.net, file = FounderHaps.xls.

Chandler J (2006) Estimating per-locus mutation rates. *J Genet Geneal*, 2:27-33.

APPENDIX

A TMRCA Estimate for Two Family Haplotypes from Scotland

To illustrate working with the modified TMRCA model, I consider a Douglas and a Hamilton haplotype; these extensive families both have roots in lowland Scotland. About four centuries ago, in fact, there was a high-level marital union between the Douglas and Hamilton families of this region, though this fact is unrelated to the examples discussed below. Both surnames have haplotypes appearing in the clade I have designated as I1-AS1 (Nordtvedt, 2008); the Hamiltons with this I1-AS1 haplotype are very numerous, and many can trace their ancestry back to either Ulster, Ireland or Lanarkshire in Scotland.

The extended haplotypes that will be used are 37-marker haplotypes as reported by Family Tree DNA, except that the markers CDYa and CDYb are ignored, leaving 35 markers to be analyzed. The Douglas and Hamilton haplotypes differ at the seven markers DYS385b, DYS439, DYS389b, DYS464d, DYS607, DYS576, and DYS570, while matching exactly on the remaining 28. The Douglas haplotype has allele values of 14, 11, 16, 16, 14, 16, 19 on those seven markers, while the Hamilton haplotype has the values 13, 12, 15, 15, 15, 17, 18 at the same markers. This results in $2^7 = 128$ first-tier possibilities for their MRCA haplotype. For the 28 markers where the haplotypes match, they also match the I1-AS1 modal haplotype, so potentially their minimum possible $D(k)$ from their clade's modal haplotype could be zero. The 128 first tier haplotypes must be inspected to find the actual minimum. But, this is straightforward since for each of the seven markers, there are just two choices for the MRCA's value, one of which in each case is equal to the I1-AS1 modal haplotype value. Therefore, we find that one of the 128 first-tier choices does, in fact, match the I1-AS1 modal haplotype, having allele values 14, 11, 16, 15, 15, 16, 19 at the seven markers, and would be the most likely MRCA haplotype. This most likely MRCA haplotype is two steps from the Douglas haplotype and five steps from Hamilton. The estimated most likely age in generations is then $7/M$, twice the traditional estimate which would be $3.5/M$ generations. The 35 markers being compared have a total mutation rate of $M = 0.111$, using the values found by Chandler (2006). For the selection of the I1-AS1 clade modal haplotype as the MRCA haplotype, the Douglas/Hamilton MRCA is estimated to have lived about 63 generations ago rather than about 32 generations ago as traditional analysis would predict. The 95%-confidence interval would range from 31 to 129 generations, again showing the very blunt nature of the TMRCA tool.