A DNA Primer

by Charles Austin

Question: Based on DNA results, what is the relationship between John Austin, Sr. and Bartholomew Austin, both of the Southside Region of Virginia? (The original question appears at the end).

I'll try to give you some helpful information, based on what I know about estimating time to MRCA (and thus estimating closeness of relationship).

If two donors have their most recent common male-line ancestor (MRCA) at most t generations back, and if they are separated from that ancestor by the same number of generations, then they are at least (t-1)th cousins (where, with t=1, "zero-th cousins" are brothers). For example, if t=2, the MRCA is either their father or their grandfather, and the donors are brothers or first cousins. If t=3, the MRCA could be their great-grandfather (3 generations back), and the donors could be as distant as second cousins.

The mathematics developed by Prof. Bruce Walsh of the Universty of Arizona (published in a 2001 paper) allows computation of the probability that two men have their MRCA within t generations, given the number (n) of markers tested, the number (k) of markers which match (so the number of mismatches is n-k) and, most critically, the mutation rates of the tested markers. Walsh first made the simplifying assumption that all markers have the same mutation rate, and for this he used .002, the average of individual marker rates as determined in two earlier studies. Later studies have suggested that this is approximately correct for the first panel of 12 markers used by FTDNA, but that the second panel (markers 13-25) has average rate about .004 while the third panel (markers 26-37) may have an average rate as high as .009. When making estimates based on the 25- or 37- marker panels I usually assume an overall average of .004 (as I have done in a couple of my other AFGS/AFAOA articles).

There is a fun and useful tool available online, called "Bill Jackson's Most Recent Common Ancestor (MRCA) Calculator" which allows the user to input n and m (the number of markers tested, and the assumed average mutation rate of the tested markers), and produce (instantly) two probability tables showing (1) the probability of MRCA within a given number of generations, and (2) the number of generations needed to achieve a given probability. The best way to find this is to go to www.genetealogy.com/resources, then click on "Genealogy Tools", and then on the link to "Bill Jackson's MRCA Calculator." If that doesn't work, let me know and I'll try to suggest something else. I really recommend that you play with this a while, trying different combinations of numbers of markers and average mutation rates. See how changing one of those variables changes the probabilities.

A footnote to the Jackson table tells you that it is based on the mathematics developed by Walsh, and gives you a reference to his paper. You can find it online and download it, if you want to see all the nasty mathematics.

This being said, Walsh also knew that more and more studies were being devoted to estimating individual mutation rates, and he gave the more complicated mathematics needed to give better probability estimates based on individual marker rates. This is what FTDNA is now doing, using rates developed by some of Walsh's Arizona colleagues led by a man with the well-known name of Mike Hammer, in a massive study using FTDNA's database of DNA profiles (haplotypes). FTDNA isn't telling what those rates are, or what their full methodology is. A number of people are trying to get their own estimated marker rates, using statistical methods on large files of DNA profiles without any prior knowledge about the mutation rates of the markers involved. I don't yet understand their methods, although I have a couple of their published papers to study. I don't really know much statistics.

So, on to your question, using the only tool I really have, the Bill Jackson Calculator.

The group in question is from Richard b. 1790 GA (donor kit 12962) through Richard d. 1759 VA (donor kit 27612), but focusing especially on the four men whose donor descendents match perfectly at 37 markers, and Bartholomew who matches them except for a mismatch (presumed due to a mutation) at marker CDYb. They are all fairly closely related, given the close DNA profiles. Even Bartholomew might not be much more distantly related to the group of four, since CDYa and CDYb are very rapidly mutating markers, with mutation rates estimated by John F. Chandler (2006) to be .03531. You read that right: only one zero after the decimal point, roughly nine times faster than the assumed average rate of .004.

The first thing I notice about the five "earliest ancestors" is that they were all probably born within the period 1750-1802, within roughly a generation of each other, in four different states: NC, VA, TN and MO. This argues strongly against their being brothers, and less strongly against being first cousins, uncle-nephew or father-son. William b. 1750 VA and John bc. 1790 VA could have the father-son or uncle-nephew relationships, however. But the MRCA of any other pair of donors in this grouping probably lies a little farther back.

All of the donors are either 4 or 5 generations removed from their earliest known male-line ancestor (that is, they are in the fifth or sixth generation beginning with that ancestor.) Bartholomew presents a puzzle. According to Mike's chart, Bartholomew had a son John b. 1806 NC and a grandson A. Jordan b. 1885 (?!) NC. Either that birthdate is wrong or a generation has been left out (or John Jordon fathered a son A. Jordan at age 79!). Carol searched some AFAOA databases, and found that Bartholomew was born 1768, that his son was named John Jordon, but that John had no son named A. Jordan or any other name with initials A.J. In fact, she couldn't find an A. Jordan Austin or A.J. Austin in the index. I'll leave you with that mystery; it's not very important to the probability estimates anyway.

Let's look at the probability that all of these men are sons or grandsons of a common ancestor one generation farther back from the oldest of them. The question, for any two of the four men (omitting Bartholomew) whose donors are perfect 37/37 matches, is: What is the probability that their MRCA is within 6 generations of the two donors. By the Bill Jackson Calculator, with n=37 and m=.004, that probability is 83.1%, which is quite high. It is probable (but of course not certain) that the MRCA of any two of these men is their father or grandfather, born about 1690-1730 in one of the southern colonies (or possibly England). The men would be brothers or first cousins or father-son or uncle-nephew. For geographical reasons noted above, I would give brotherhood a much lower probability, however. But that's a problem for the paper chase.The common MRCA of all four of them might be a generation or two farther back.

Now, what about Bartholomew? Comparing him with any of the other four, the donors match 36/37. If we assume (as the Jackson tables do) that all markers have mutation rate .004, the table gives probability of only 52.3% that their MRCA occurred within 6 generations from the donors. But wait! The actual mutation rate of the guilty marker CDYb is about .035, almost nine times higher. A mutation at that marker is about 9 times more likely to occur in one change of generation than at one of the "average" markers. So, a mutation there has less adverse impact on the probability than would one at an "average" marker. The true probability that the MRCA of the donors of Bartholomew and one of the others is at most 6 generations back is probably between 52.3% and 83.1%, and probably closer to the higher figure. But you asked about the relationship between Bartholomew and "John of Southside," b.1992, VA. That John, unfortunately, has been tested only at 25 markers, and his donor apparently has two mismatches with the Bartholomew donor within those first 25 markers. A 23/25 match doesn't give a very high probability of an MRCA within 6 generations, or even 8 or 10.

But (again) wait! The markers 389-1 and 389-2 are actually overlapping pieces of DNA; 389-1 is really the first part of the string forming 389-2. Both markers have the same "repeat motif" (which is either

TCTA or TCTG). The Bartholomew donor has 14 repetitions of the motif at 389-1, while the John donor has 15. That's the mutation. The rest of 389-2 has 17 repetitions in both donors, giving total lengths of 31 and 32 respectively. (That's why the 32 on Mike's chart isn't shaded: it doesn't represent a second mutation. So we really have a 24/25 match here.) Here are the Bill Jackson results for a 24/25 match with average mutation rate .004: The probability that the two donors have their MRCA within 6 generations is 32.9%. (If it were a 23/25 match, it would be only 11%). The number of generations (from the donors) needed to get at least a 50% probability is 8.7 (about 260 years, at 30 years/generation). To get a probability of 75%, you have to go back 13.9 generations (417 years) from the donors. So, on the basis of present knowledge, the relationship between Bartholomew and this John could be that of relatively distant cousins, although a closer relationship is possible. If John were extended to 37 markers, you might get a better result, if the last 12 markers matched 12/12 or at least 11/12

Jim, I hope this all is helpful to Liz and Jane, and I hope you'll give yourself the fun of bringing up Bill Jackson at least to confirm the results I've given you.

Referring back to your note, I think I would say that the four whose donors match 37/37 are "very closely" related to each other, probably first cousins or even brothers; that Bartholomew is probably a first or second cousin to them and hence also "very closely" related to them, but that Bartholomew and John of Southside might be only "somewhat closely" related.

There is a question of terminology you should be careful about. The kits represent the DNA of the donors, not of their distant ancestors. I often start out talking about two donors, their DNA matches and the closeness of their relationship, and soon find that I have forgotten about the donors and am talking about the DNA and relatedness of the donors' ancestors instead. As in "William, John, Joseph and John are perfect matches." They probably are, but it's the donors' DNA we're testing. We can't look at the ancestors' DNA (unless we dig them up).

There is a common assumption that if two men have the same DNA profile, so does their MRCA. I came to believe that the probability of there having been no mutations in either line of descent from the MRCA to the matching donors is very high. However there are two unlikely events in which a pair of mutations could result in perfectly matching donors: (1) two mutations in opposite directions at a single marker, leaving it in its original state, or(2) the same mutation at the same marker in each line of descent (so-called "parallel mutations"). I calculated very low probabilities for such events, leaving a very high probability that there have been no mutations at all. Note that in the case of parallel mutations, the matching donors are not a perfect match for the MRCA: they differ from the MRCA at that one marker. Unlikely as it is, there is a case of parallel mutations in Mike's table, at DYS 576. It occurs in the Samuel line.

----- Original Question -----

We are exploring the relationship of Bartholomew Austin to John Austin of Southside. At issue here is how to interpret DNA test results. How many marker mismatches in which position disqualifies one from stating beyond a doubt, that one person is related to another?

Looking at Mike's DNA chart, (please see kit # 32962, 129275, 1812, etc.) perfect 37 marker matches occur for William, John, Joseph, & John. Bartholomew differs only by 1 marker out of 37, and it is out in position 35, certainly a good enough match to these others.

Below these people, but in the same grouping is John (b. 1692) of Southside with 2 mismatches, Nathaniel with 2 (?) mismatches and Richard with one mismatch.

We are stating that Bartholomew is only very closely related to John. Are we being too strict in our interpretation?

Thank you.