

Journal: www.jogg.info

Originally Published: Volume 3, Number 2 (Fall 2007)

Reference Number: 32.003

LETTERS TO THE EDITOR (REGARDING 'GEOGRAPHIC PATTERNS OF HAPLOGROUP R1B IN THE BRITISH ISLES' FROM VOL. 3, NUM. 2 (SPRING 2007))

Author(s): R.H. McGregor, responses by Kevin G. Campbell

Letters to the Editor

Picts and Gaels

To the Editor:

In the article, "Geographic Patterns of Haplogroup R1b in the British Isles," which appeared in the Spring 2007 issue of JoGG, Kevin Campbell asserts that "OGAP4 best represents the Pictish ancestry of Scotland". He further asserts that, "The Gaels of Ireland, as identified by the DNA signature of OGAP8, are as close as any group to being considered the root line and forbearers of Celts today. When present in Scotland, it is suggested that OGAP8 represents the signature of the Dal Riada Celts."

I disagree with these conclusions. Using the principles of convergence analysis, diversity (genetic distance) and founders modal values, I have analyzed five sets of R1b data and developed the modal values for each set over 37 DYS loci . . . [and identify the Picts and Dal Riada Celts differently].

R. H. McGregor

[Ed Note: Mr. McGregor presented a detailed analysis of why he disagreed with the statements that he quoted above. However, as the response from Mr. Campbell shows, this analysis is beyond the scope of his article (and the Letters to the Editor feature) and so it was not necessary to include it here in order for Mr. McGregor's main comments to be addressed]

Campbell Responds:

First and foremost, I would like to make it clear that the purpose of my article was not to develop new truths regarding the Celts and the Picts. My aim was to attempt to infer Prof. Sykes' DNA definitions of them. Sykes devotes a chapter of his book to the Picts, and one commercial lab even offers a Pict test. Obviously, some geneticists believe that they have an idea of the DNA signature that defines Pict-ishness.

I believe I accomplished my objectives in the article. I re-analyzed the OGAP data and found reasonable evidence that OGAP8 and OGAP4 are the defining haplotypes for the clans identified by Sykes as Dal Riada Celts and Picts. Whether Sykes' identifications are ultimately correct or not, is another question.

Randomness of Mutations

To the Editor:

In the article, "A Major Subclade of Haplogroup G2," by T. Whit Athey (Spring 2007), in the discussion of Table 3, Athey states: "Table 3 illustrates the ratios of the variance in the two populations on each of the 29 DYS markers. Because of the random nature of mutations, the following ratios" I disagree with this statement. I do not believe that STR mutations are random. I will cite two published papers and some of my convergence analysis results to show that mutations are in some sense "controlled/constrained."

First, I will cite a Review paper: "Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference", D.B. Goldstein and D.D. Pollock, *Journal of Heredity*, 88:335-342, 1997. In discussing "Range Constraints on DYS Loci" mutations the authors state: "Perhaps the most compelling evidence that the number of repeats at microsatellite loci is under some form of constraint is simply the absence of alleles of very large size. Given the high mutation rate, and the very large number of loci that have been characterized, it is clear that if the process were an unconstrained random process we would expect to regularly observe loci with very large alleles." Note that subsequent databases published on DYS loci values for different Haplogroups support this observation.

Second, consider the paper: "Genealogical and Evolutionary Inference with the Human Y Chromosome, M.P. Stumpf and D. B. Goldstein, *Science*, 291:1738-1742, 2001. In this article the authors state: "The expected value [of the average squared distance], averaged over all alleles is thus an unbiased estimator for TMRCA. In practice the equation would be evaluated for each of many loci and averaged." The importance of this statement is the assertion that any DYS loci can be used to estimate TMRCA, if more than one is used, then the average over all DYS loci is estimated for TMRCA, i.e. each DYS loci is an equal contributor to TMRCA!

Third, I have analyzed the non-iberian Tarin data set. Using all DYS loci I find that the TMRCA is 7352 BP for this set of entries. Further, if I use individual dys loci I get the following TMRCA's: 393: 7398; 391:7406, 389ii: 7325; CDYa: 7405; CDYb: 7414; 449: 7382. The range of these estimates is within 0.7% of the mean. This confirms the results of the second

reference. To me, the implications of these results are quite clear. DYS loci mutations are “constrained” in some manner.

To carry this discussion much further becomes an issue of philosophy more than genetic calculations, however it seems clear to me that mutations are not random!

R. H. McGregor

Whit Athey Responds:

In regard to Mr. McGregor’s first point, he has misunderstood what I meant when I referred to the “randomness” of mutations. I was not referring to randomness of the length of STR markers, but to the randomness of when mutations occur in time. I certainly agree that several articles have presented compelling evidence that the lengths are constrained. There is also evidence that the mutation rate is inversely correlated with length. However, none of this is relevant to my statement.

In regard to the second point, it is not clear why Mr. McGregor finds it necessary to make this comment in

regard to my article. My approach is entirely consistent with the statement Mr. McGregor quotes. In fact, I averaged the TMRCA over all the markers that I had available to me, just as the cited reference recommends. I fail to see any disagreement.

In regard to the third point, Mr. McGregor appears to be a victim of circular reasoning. He has first calculated the TMCRA for the dataset using Zhivotovsky’s average rate over seven markers and the ASD for the seven markers. Then he used that TMRCA and the ASD for each individual marker to calculate mutation rates for each of the 37 markers. This much, in principle, is valid (at least the results are as good as the dataset). But, he then used those individual mutation rates to recalculate the TMRCA for each marker and found that he got back almost exactly the same TMRCA on each marker that he got for the whole dataset. However, those individual TMRCA’s he calculated are not just approximately the same as the TMRCA for the dataset, they are IDENTICALLY equal to it! The fact that he gets the same TMRCA for each marker has no significance at all—it had to come out that way. It does not confirm his second point at all, but his second point is well accepted by everyone and luckily doesn’t need confirming.