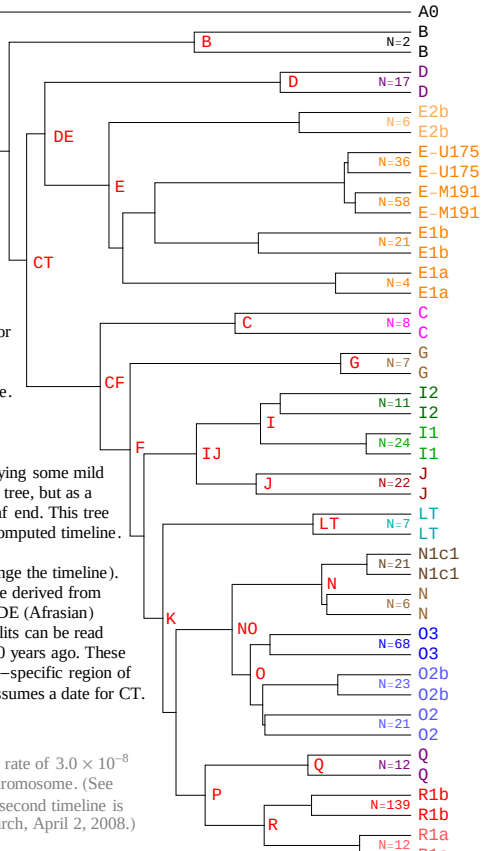
y−Haplogroup Tree

Y

Based on 18,692 variable sites and 526 people extracted from the 1000 Genomes Project.



There are around 24 million nucleotides in the euchromatic male−specific region of the human Y chromosome. On average, only about one of those nucleotides will mutate when passed from father to son. If you were to know the entire nucleotide sequence for the Y chromosome of one individual and compare that with another individual, you could, if you knew the Y chromosome nucleotide mutation rate, estimate the number of generations separating the two individuals. You simply count the number of nucleotide differences between the two Y chromosomes and use a mutation rate to convert to time.

Such an SNP counting method for estimating the time to most−recent−common−ancestor (TMRCA) of two individuals, would be independent of any STR marker based method.

The method would require much of the Y chromosome to be sequenced to make it viable. To test the method, I extracted some relevant data from the 1000 Genomes Project – in particular, 526 males with Y chromosome coverage as at 2012−04−03.

By simply counting the nucleotide differences between those Y chromosomes, and applying some mild priors, the tree shown here can be constructed. All 526 males were used to construct the tree, but as a simplification only a count of those people, and their y−Haplogroup, is shown at the leaf end. This tree is equivalent to the Y−DNA phylogenetic tree, but with the added feature of having a computed timeline.

The first timeline is determined by assuming a mutation rate† (change the rate, then change the timeline). According to that computed timeline, the Y−chromosome of all the 526 men would have derived from a single male who lived approximately 140,000 years ago. The major split between the DE (Afrasian) and CF (Eurasian) haplogroups would have occurred around 55,000 years ago. Other splits can be read from the timeline, for instance the I1 and I2 haplogroups would have split around 21,000 years ago. These estimates are all based on comparing the nucleotide differences in the euchromatic male−specific region of the human Y chromosome, and using the assumed mutation rate. The second timeline assumes a date for CT.

† The first timeline, below, is an estimate based on using an assumed germline mutation rate of 3.0×10^{-8} mutations per nucleotide per generation for the male−specific region of the human Y chromosome. (See Cruciani et al, American Journal of Human Genetics 88, 814–818, June 10, 2011.) The second timeline is calibrated by arbitrarily fixing an age for the CT split. (See Karafet et al, Genome Research, April 2, 2008.)

Calibration so that Y splits at 140 kya.

140 kya | 120 kya | 100 kya | 80 kya | 60 kya | 40 kya | 20 kya | 0 kya

Calibration so that CT splits at 70 kya.

160 kya | 140 kya | 120 kya | 100 kya | 80 kya | 60 kya | 40 kya | 20 kya | 0 kya