

Non-Paternity and Locus Specific Mutation Rates of 36 Y Chromosome STRs

K. Hadley^{1,2}, N.M. Myres^{1,2}, J.B. Ekins^{1,2}, J.E. Ekins^{1,2}, L.A.D. Hutchinson^{1,2}, D. Layton^{1,2}, M.L. Lunt^{1,2}, S.S. Masek^{1,2}, A.A. Nelson^{1,2}, M.E. Nelson^{1,2}, K.L. Pennington^{1,2}, U.A. Perego^{1,2}, J.L. Peterson^{1,2}, A. Sims^{1,2}, T. Tölley^{1,2}, A. Welch^{1,2}, S. Woodward^{1,2} 1) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT; 2) Computer Science, Brigham Young University, Provo, UT; 3) Department of Micro and Molecular Biology, Brigham Young University, Provo, UT.

Abstract

Mutations and non-paternities are two of the most influential factors that may bias conclusions in Y-chromosome (y-ch) population genetic and lineage studies. Previous studies have determined locus specific mutation rates for 12 y-ch loci. Herein, we characterize mutation rates of 36 STR loci based on 500 individuals separated by between 1-13 meioses, including eight previously reported loci for comparison to published rates. Non-paternity rates have been shown to vary among certain demographics, age groups, and time periods. To account for all of these factors, the present study reports a non-paternity rate for samples collected from over 100 different families that vary in age group, culture, country of origin and the number of meioses separating them.

Complete characterization of several Y-chromosome loci is an important task, the results of which impact a variety of research areas: evolution, speciation, forensics, genealogy, and others. A fundamental property of a Y-chromosome locus is its mutation rate. Individual locus mutation rate estimates are important to accurately calculate relatedness between individuals in populations. Previous studies have determined locus specific mutation rates for 12 Y-ch loci (Kayser 2000, Heyer 1997). Heyer et al calculated mutation rates using deep rooted pedigrees, while Kayser et al calculated rates based on father/son relationships. We report mutation rate estimates calculated from both deep-rooted and father/son pedigrees.

Mutation rates for 36 Short Tandem Repeat (STR) loci were calculated based on 135 multigenerational pedigrees and 73 father/son pairs (see Table 2). Individual samples represent 18 world populations, predominantly originating from European countries and the continental United States.

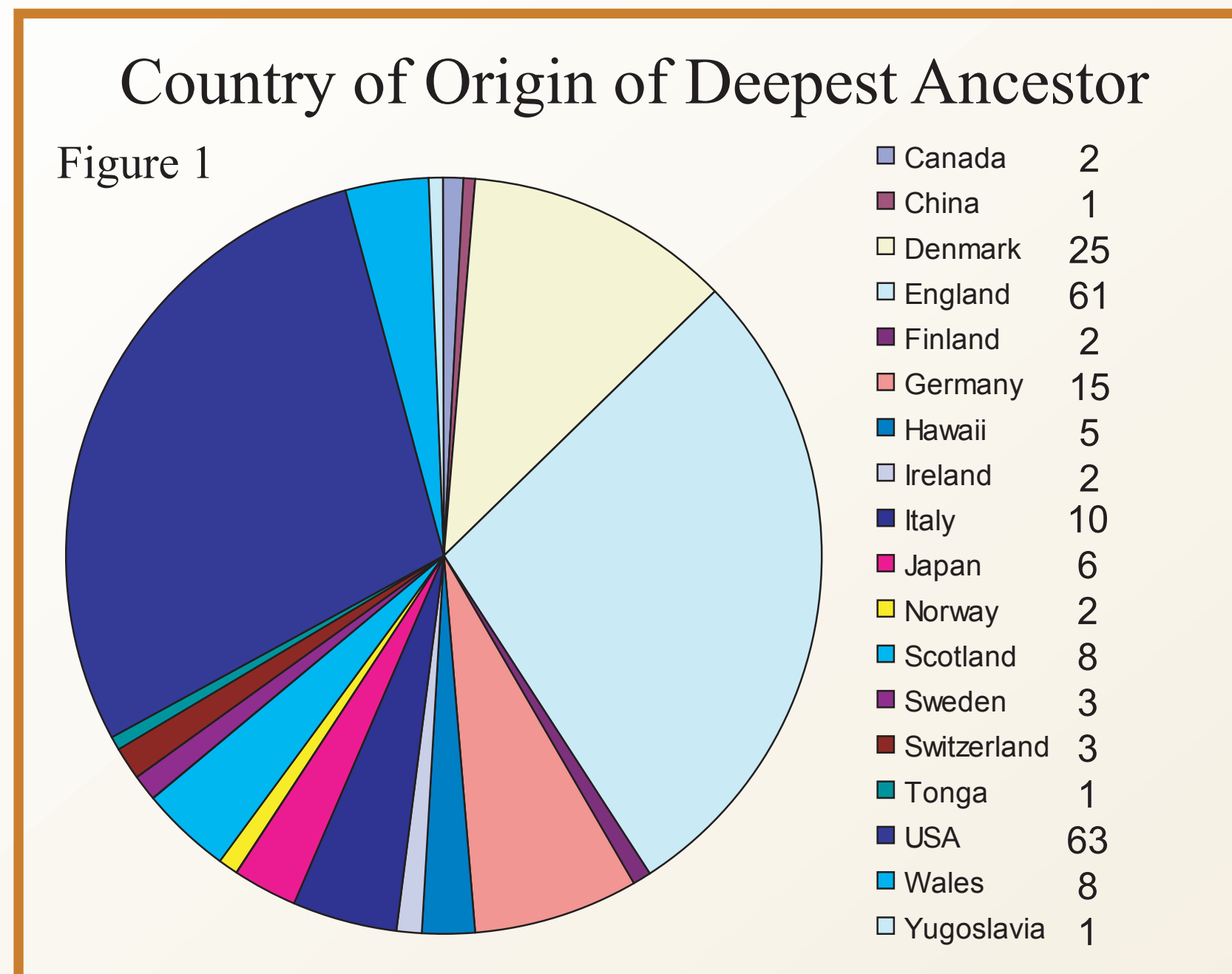
Mutation Rates

Non-Paternity Rates

Accurate non-paternity rates are critical for population and family based studies and for the accurate reconstruction of recent genealogies. Systematic studies of non-paternity rates are limited (M. A. Jobling 2004). Early reports by MacIntyre and Sooman (1991) report rates between 1 to 20%. More recently a study in Switzerland found a rate >1% (Sasse 1994). A study from Mexico established a rate of 12% (Cerde-Flores 1999). Several factors that increase the difficulty in determining an accurate non-paternity rate include variations in demographics, culture, and age groups (Cerde-Flores 1999, Jobling 2004, MacIntyre and Sooman 1991). It is also suggested that rates will vary over different time periods, with the rate decreasing back in time. Recently AABB accredited laboratories reported 28.70% of their cases as non-paternities (AABB Annual Report Summary for Testing in 2002). However, it is expected that this rate is considerably higher than the non-paternity rate in the general population due to sampling bias.

In the present study, the derived rates result from samples collected from families that vary in age group, culture, country of origin (see Figure 1) and the number of meioses separating them (see Table 1). It is presumed that the resulting rates represent an average for all of these factors. In addition, no selection for suspected non-paternity has occurred. To preclude the necessity of obtaining a sample from both parents, Y-chromosome (Y-ch) testing was applied in determining non-paternity rates in purported paternal lines.

Pedigree information was collected from each participant. The paternal line was extended using the Ancestral File database (see Family Search website). Only individuals without known non-paternal event were included in our analysis. Individuals shown to share a common ancestor along their strict paternal line were grouped into Common Lines (CL) and the number of meiotic events were counted. Equation 19b (Walsh 2001) was used to calculate the Most Recent Common Ancestor (MRCA) for each CL using the infinite alleles model and individual mutation rates (not based on pedigrees) from Hutchison (in prep). The cumulative likelihood (cum) was calculated for the expected CA that was indicated by the genealogy (see Table 1). This was converted to an Odds Ratio: $OR = (1 - cum) / cum$. The odds ratio is the likelihood of a more distant vs. more recent CA than indicated by the genealogy. For example CL 77 has two mutations in two meiotic events. The cumulative likelihood is 0.007 which converts to odds of 146:1 that there is a more distant common ancestor than 2 meioses separating these individuals. The cut off point of 35:1 was determined by a threshold between 1 and 2 mutations in 1 to 2 meioses. Observing two mutations in two meioses is unlikely and the CL was determined to have a non-paternity. The non-paternity rate was calculated by dividing the number of pairs with a non-paternity by the average number of meiotic events between all pairs of individuals. The rate was determined to be, 1.27%. For our calculations, we assumed only one non-paternity event per non-paternity pair. It is also assumed that incorrect genealogical data, either from Ancestral File or mistakes submitted from the participant are incorporated in our non-paternity rate. It is difficult to determine the exact generation in which a non-paternity occurs, therefore this frequency reflects an average over all generations separating the samples.



Conclusions

An accurate non-paternity rate is important for population studies, recombination estimates, and estimating transmission probabilities of genetic disorders (Cerde-Flores 1999). Here we report a non-paternity rate that is lower than the previous estimate from Cerde-Flores et al. However, some recent studies have also detected lower rates, such as the Sasse et al (>1%) and Sykes 2000 (1.3%/generation) studies. Our rate was calculated based on samples originating in 18 world populations, predominately from European countries and the continental United States.

Individuals that had known non-paternities, or adoptions in their pedigree were excluded from this study. Thus, the 1.27% proposed rate describes a non-paternity rate for individuals that have no knowledge of a non-paternity event in their ancestry and it may be lower than that of the general population. These rates were calculated based on multigenerational pedigrees and 73 father/son pairs.

We report locus specific mutation rates for 36 Y-ch STR loci, the largest number of loci to date. We compared our mutation rates to those previously published by Heyer et al and Kayser et al. Our rate estimates are comparable to the 9 loci previously published. We estimate rates for an additional 27 loci not previously published.

Table 2 Individual Locus Mutation Rates

Locus	# of Mutations	# of Meioses	# of families	Mutation rates	Heyer (213 gens)*	Kayser (btwn 240-996 gens)
DYS385 **	3	888	206	0.003378	0.0047	0.0021
DYS388	0	888	206	0	-	-
DYS389	1	888	206	0.001126	0.0094	0.00235
DYS389II	1	888	206	0.001126	-	0.00471
DYS390	3	888	206	0.003378	0	0.00858
DYS391	1	888	206	0.001126	0	0.00482
DYS392	2	888	206	0.002252	0.0047	0
DYS393	2	889	207	0.00225	0	0
DYS19/394	0	889	207	0	0	0.00201
DYS426	0	858	199	0	-	-
DYS437	0	889	207	0	-	-
DYS438	0	888	206	0	-	-
DYS439	6	888	206	0.006757	-	-
DYS441	1	820	190	0.00122	-	-
DYS442	4	816	188	0.004902	-	-
DYS444	1	805	186	0.001242	-	-
DYS445	2	845	196	0.002367	-	-
DYS446	6	799	182	0.007509	-	-
DYS447	2	858	199	0.002331	-	-
DYS448	1	888	206	0.001126	-	-
DYS449	11	889	207	0.012373	-	-
DYS452	0	664	151	0	-	-
DYS454	0	889	207	0	-	-
DYS455	0	851	198	0	-	-
DYS456	7	845	196	0.008284	-	-
DYS458	8	889	207	0.008999	-	-
DYS459 **	0	889	207	0	-	-
DYS460	2	878	205	0.002278	-	-
DYS461	0	889	207	0	-	-
DYS462	1	889	207	0.001125	-	-
DYS463	2	800	186	0.0025	-	-
GGAAT1B07	1	889	207	0.001125	-	-
YCAII	0	858	199	0	-	0.00204
YGATAA10	1	889	207	0.001125	-	-
YGATAC4	4	889	207	0.004499	-	-
YGATAH4	3	888	206	0.003378	-	-
average	2.111111	864.806	200.6944	0.002438	0.0021	0.0028

* Heyer's most conservative mutation rate in his 1997 paper.

** combined values for the mutation rate and gene diversity are given for DYS385, DYS459, and YCAII because they are duplicated loci and can not be distinguished.

Table 1 Determination of Non-paternities

Common Line	Mutations in CL	# Missing loci	# of Meioses	Cumulative Likelihood	Odds Ratio*
174	0	0	13	0.8944387	0.11802
209	0	0	13	0.8944387	0.11802
104	0	0	12	0.8759447	0.141625
158	0	0	12	0.8759447	0.141625
192	0	0	12	0.8759447	0.141625
157	0	0	11	0.8542107	0.170671
173	0	0	11	0.8542107	0.170671
150	0	0	10	0.828669	0.206754
151	0	0	10	0.828669	0.206754
34	0	0	9	0.7986525	0.252109
180	0	0	9	0.7986525	0.252109
228	0	0	9	0.7986525	0.252109
168	0	0	8	0.7633771	0.309968
183	0	0	8	0.7633771	0.309968
197	0	0	8	0.7633771	0.309968
205	0	0	8	0.7633771	0.309968
214	0	0	8	0.7633771	0.309968
218	0	0	8	0.7633771	0.309968
222	0	0	8	0.7633771	0.309968
226	0	0	8	0.7633771	0.309968
18	0	7	10	0.7536032	0.326958
21	0	2	8	0.7436718	0.344679
164	0	0	7	0.7219217	0.385192
167	0	0	7	0.7219217	0.385192
211	0	0	7	0.7219217	0.385192
156	0	0	6	0.6732035	0.485435
186	0	0	6	0.6732035	0.485435
199	0	0	6	0.6732035	0.485435
202	0	0	6	0.6732035	0.485435
20	0	1	6	0.666108	0.501258
85	0	1	6	0.666108	0.501258
54	0	3	6	0.6452275	0.549841
8	0	6	7	0.6449887	0.550463
74	0	4	6	0.6260594	0.597292
96	0	5	6	0.6162723	0.622659
133	0	0	5	0.61595	0.623508
165	0	0	5	0.61595	0.623508
203	0	0	5	0.61595	0.623508
195	2	1	20	0.6153147	0.625195
82	0	2	5	0.6088373	0.642475
129	0	0	4	0.548666	0.822603
130	0	0	4	0.548666	0.822603
137	0	0	4	0.548666	0.822603
177	0	0	4	0.548666	0.822603
194	0	0	4	0.548666	0.822603
38	0	1	4	0.5417432	0.845893
204	0	1	4	0.5417432	0.845893
79	1	4	11	0.5081756	0.967824
40	0	0	3	0.469594	1.129499
127	0	0	3	0.469594	1.129499
132	0	0	3	0.469594	1.129499
136	0	0	3	0.469594	1.129499
189	0	0	3	0.469594	1.129499
193	0	0	3	0.469594	1.129499
196	0	0	3	0.469594	1.129499
14	0	1	3	0.4631409	1.15917
80	0	2	3	0.4581655	1.182618
36	1	0	9	0.4502378	1.221049
208	1	0	9	0.4395649	1.274977
27	1	1	9	0.4364332	1.291301
100	0	7	3	0.4145905	1.412019
122	1	0	8	0.3960654	1.524836
153	1	0	8	0.3929477	1.544868
99	1	0	8	0.3895538	1.567039
152	1	0	8	0.3893057	1.575729
191	1	0	9	0.3883057	1.575729
44	1	1	8	0.378632	1.641087
109	0	0	2	0.3766889	1.654852
124	0	0	2	0.3766889	1.654852
139	0	0	2	0.3766889	1.654852
140	0	0	2	0.3766889	1.654852
144	0	0	2	0.3766889	1.654852
155	0	0	2	0.3766889	1.654852
169	0	0	2	0.3766889	1.654852
180	0	0	2	0.3766889	1.654852
200	0	0	2	0.3766889	1.654852
210	0	0	2	0.3766889	1.654852
213	0	0	2	0.3766889	1.654852
217	0	0	2	0.3766889	1.654852
227	0	0	2	0.3766889	1.654852
23	0	1	2	0.3710563	1.695009
81	0	1	2	0.3710563	1.695009
93	0	12	3	0.3621041	1.761636
19	0	2	2	0.3607783	1.771786
1	0	3	2	0.3509815	1.849153
170	2	0	13	0.3489513	1.882249
159	1	0	7	0.3440949	1.906175
52	0	4	2	0.3405652	1.936296
83	0	4	2	0.3405652	1.936296
47	1	0	7	0.3398507	1.942468
121	1	0	7	0.3398507	1.942468
3	1	5	8	0.33789	1.959543
212	1	0	7	0.3311749	2.019553
94	1	1	7	0.3311743	2.019558
57	1	1	7	0.3274299	2.054089
58	1	1	7	0.3224635	2.101126
76	1	2	7	0.3175781	2.148832
48	0	2	7	0.3175614	2.148997
123	1	0	6	0.2841458	2.51932
73	1	0	6	0.2836889	2.525237
134	1	0	6	0.2825382	2.539344
118	1	4	7	0.2779984	2.597143
2	0	0	1	0.2674637	2.738825
6	0	0	1	0.2674637	2.738825
30	0	0	1	0.2674637	2.738825
35	0	0	1	0.2674637	2.738825

* Odds Ratio of more distant Vs. more recent C.A. indicated by the genealogy.

References

Cerde-Flores, R., S. Barton, L. Marty-Gonzalez, F. Rivas, R. Chakraborty. 1999. Estimation of Nonpaternity in the Mexican Population of Nuevo Leon: A Validation Study with Blood Group Markers. *Am J Phys Anthropol.* 109:281-293.

Heyer, E., J. Puymirat, P. Dieltjes, E. Bakker, P. de Kniff. 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.* 6:799-803.

Jobling, M. A., M. Hurles, C. Tyler-Smith. Human Evolutionary Genetics Origins, Peoples & Disease. New York: Garland Science Publishing. 2004.

Kayser, M., L. Roewer, M. Hedman, (11 co-Authors). 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 66:1580-1588.

MacIntyre S., A. Sooman. 1991. Non-paternity and prenatal genetic screening. *Lancet* 338:869-871.

Parentage Testing Program Unit. AABB Annual Report Summary for Testing in 2002. Nov 2003. http://www.aabb.org/About_the_AABB/Stds_and_Accred/ptannrpt02.pdf. (25 Oct. 2004).

Sasse, G., S. Müller, R. Chakraborty, J. Ott. 1994. Estimating the Frequency of Nonpaternity in Switzerland. *Hum Hered.* 44:337-343.

Sykes, B., C. Irven. 2000. Surnames and the Y Chromosome. *Am J Hum Genet.* 66:1417-1419.

Walsh, B., June 2001. Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics* 158, 897-912.

Electronic Database Information: Family Search, <http://www.familysearch.org> (for Ancestral File)