

SHORT REPORT

The HNPCC associated *MSH2**1906G→C founder mutation probably originated between 1440 CE and 1715 CE in the Ashkenazi Jewish population

S Sun, C M T Greenwood, I Thiffault, N Hamel, G Chong, W D Foulkes

J Med Genet 2005;42:766–768. doi: 10.1136/jmg.2005.030999

The *MSH2**1906G→C mutation was recently shown to be a rare yet highly penetrant mutation leading to colorectal cancer. The mutation was only found among Ashkenazi Jewish individuals and lies on an extended haplotype that is common in that population. This study determined that the mutation probably arose between 11 and 22 generations ago, during the time when the Ashkenazim were living in eastern Europe.

In both men and women, colorectal cancer is the third most commonly diagnosed cancer in Canada.¹ A small proportion of colorectal cancer cases (less than 5% overall) can be attributed to highly penetrant mutations in susceptibility genes such as *APC*, *MLH1*, and *MSH2*.²

Recently, we reported identification and characterization of a new mutation in *MSH2* (*MSH2**1906G→C) among Ashkenazi Jewish individuals with colorectal cancer.³ The mutation appears to be rare in non-Ashkenazi Jews and is highly penetrant. Among 16 mutation carriers studied, recruited from five different countries, all carried the same haplotype around the mutation, indicating that all cases probably descend from a common ancestor. It seems clear that this mutation can be considered as a founder mutation in this population⁴; we have therefore estimated the probable age of the mutation in order to interpret its frequency and distribution in light of the demographic history of the Ashkenazim.

The usual method for estimating mutation age⁵ is based on the expected decay of linkage disequilibrium owing to recombination over the generations between time of the most recent common ancestor and today. Modifications have been suggested that adjust for population growth under a simple model.⁶ However, consideration of gene genealogies and models for coalescent times for different chromosomal lineages can lead to richer models capable of accounting for the randomness of the gene genealogy.

An intra-allelic coalescent model for a rare disease mutation, allowing for population growth and selection, was recently extended to multipoint linkage disequilibrium mapping.⁷ This approach elegantly incorporates the dependence between gene genealogies at multiple markers. Incorporation of realistic assumptions about the past leads to higher variability in the genealogies, and therefore the confidence regions for mutation age tend to be wider than regions from simpler approaches. It is also worth noting that this method estimates the time of origin of the mutation, and this may be substantially older than the time of the most recent common ancestor, which is the quantity estimated by other methods.^{3–7} Software is available (DMLE).⁸

Order and physical distances between markers were updated (UCSC, Santa Cruz, California, USA, July 2003).

Genetic distances were taken from deCODE⁹ where available. Linear interpolation was used to obtain genetic distances for markers not in deCODE and for one marker (D2S391) where the genetic distance given by deCODE⁹ gave an incorrect marker ordering. The single locus method requires knowledge of haplotype phase. For the 12 families where phase was unknown, it was estimated using the software PHASE¹⁰; we seeded the algorithm with the haplotypes of the four phase-known cases, and sampled from the posterior distribution of mutation carrying haplotypes. We estimated population growth among Ashkenazi people from the year 1500 onwards at 1.5-fold or 1.6-fold per generation, assuming 25 years per generation.⁵ The proportion of mutation carrying chromosomes sampled was estimated from a lifetime risk of colorectal cancer of 6.85%,¹ a worldwide population of Ashkenazi Jews of 13 million,⁴ and a mutation prevalence among colorectal cancer cases of 0.59%.³

The single locus method for estimation of the time since the founding of the mutation^{5,6} gives highly variable estimates ranging between 12 and 700 generations (table 1). In particular, for markers near the mutation the estimates are very sensitive to the recombination rates and also may be strongly affected by mutation at the microsatellite markers. However, the coalescent model with growth rate 1.5-fold per generation (fig 1A) presents a clearer picture of a recent mutation origin at about 17 generations ago (95% confidence region, 12.7 to 22.5 generations). For a growth rate of 1.6, the mutation is estimated to be slightly younger (fig 1B).

Estimates of the year of origin of the mutation range between 1440 (growth rate 1.5-fold per generation, upper end of the confidence region) and 1715 (growth rate 1.6-fold per generation, lower end of the confidence region). This 275 year range places the mutation in the period when the Ashkenazim were living in northern Europe, in partially closed communities. Such inference makes sense, as this mutation has not been identified in other Jewish groups or in individuals not of Jewish descent. These dates are consistent with an origin around the time of the Chmielnicki massacres that particularly affected the Jewish population of what is now Ukraine and Poland in 1648 to 1655. A similarly recent age and site of origin has been postulated for the mutation in the torsion dystonia gene, although that mutation is frequent in the Ashkenazim.⁵ For this rare *MSH2* mutation, the rapid population growth means in fact that neither the existence of positive selection nor drift are required to explain the current allele frequency.

The estimate of lifetime colorectal cancer risk (6.85%) was taken from Canadian cancer statistics for 2004. This rate may be somewhat higher than estimates from the USA or from Canada several years previously. We repeated the DMLE analysis using a lifetime risk of 5% for colorectal cancer, and

Abbreviations: DMLE, disease mapping using linkage disequilibrium

Table 1 Associated allele frequencies, linkage disequilibrium, and single locus estimates of the number of generations since the most recent common ancestor

Marker	Associated allele†	θ	P_n	P_d	δ	\hat{g}	Labuda corrections*	
							$r=1.60$	$r=1.50$
D2S2331	4	0.077	0.22	0.49	0.35	13.25	3.36	3.61
D2S2306	4	0.045	0.35	0.82	0.72	7.18	4.50	4.93
Rs868542	C	0.021	0.70	0.68	-0.07	NA		
D2S391	2	0.011	0.19	0.69	0.61	42.40	7.42	8.31
Rs1374749	A	0.0098	0.52	0.77	0.51	67.33	7.75	8.69
Rs896213	G	0.0040	0.38	0.89	0.81	51.11	9.65	10.89
D2S2227	3	0.00039	0.28	0.82	0.76	723.03	14.63	16.67
MSH2*1906G→C	-	0.0	0.0	1.0	NA	NA		
CA3	139	0.00088	0.15	0.83	0.80	256.79	12.89	14.65
CA1	143	0.00090	0.13	0.82	0.79	256.33	12.84	14.59
CA2	250	0.0011	0.54	0.94	0.86	136.55	12.41	14.10
D2S123	6	0.049	0.48	0.60	0.23	29.11	4.32	4.71

Risch *et al*⁸ estimated mutation age using generations $\hat{g} = 1n(\delta)/1n(1-\theta)$, where θ is the recombination fraction between the marker and the mutation, P_n and P_d are the frequencies of the associated marker allele on normal and mutation carrying chromosomes, respectively, and linkage disequilibrium between the marker and the mutation is measured by $\delta = (P_d - P_n)/(1 - P_n)$. Twenty five years per generation was assumed, and r denotes the growth rate per generation.

*Labuda corrections are added to the single locus estimates (\hat{g}) to obtain unbiased estimates of mutation age under the assumptions of a bifurcating genealogy.

†See Foulkes *et al*⁹ for details of allele numbering.

this did not alter our estimates of mutation age detectably. On the other hand, the coalescent model is very sensitive to the assumed growth rate. However, for this population there is good information about the population size over the last 500 years. Nevertheless, variation in population size owing to persecution or major epidemics was not considered, and such factors could affect the variability of gene genealogies and hence the estimated mutation age. The effect of the massacres in eastern Europe is likely to increase dependence between gene genealogies, and therefore to make the age of the mutation appear slightly older; on the other hand, any selection against the mutation would make the mutation appear slightly younger. Selection is not likely, given that most cases have onset of cancer in the fourth or fifth decades of life.

This mutation was identified by sequencing a well studied candidate gene, *MSH2*, in which several other mutations have also been shown to increase risk for colorectal cancer.² It is worth speculating whether a strategy of testing for linkage disequilibrium could have identified this mutation. Recently, it has been shown that genome-wide association testing was able to identify this gene using only Ashkenazi Jewish individuals with colorectal cancer (compared with Ashkenazi controls).¹¹ Therefore, case-control studies of rare variants can be a successful strategy when cases come from high

risk families, when a founder population with extensive linkage disequilibrium is studied, when the founding mutation is recent enough for the haplotypic signature to extend over a sizeable region, and when there is one mutation responsible for a large proportion of the cases in the sampled population.

ACKNOWLEDGEMENTS

This work was supported by the Mathematics of Information Technology and Complex Systems/Networks of Centres of Excellence Program, and by the Ontario Genomics Institute. WDF is funded by the Canadian Genetic Diseases Network and the Cancer Research Society Inc. We appreciate Dr Reeve's assistance and suggestions with respect to the DMLE software, and Dr Speed for his comments.

Authors' affiliations

S Sun, Department of Statistics, University of Toronto, Ontario, Canada

C M T Greenwood, Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto

I Thiffault, N Hamel, Division of Medical Genetics, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

G Chong, Department of Diagnostic Medicine, Sir M B Davis-Jewish General Hospital, McGill University, Montreal

W D Foulkes, Program in Cancer Genetics, McGill University, Montreal, Quebec, Canada

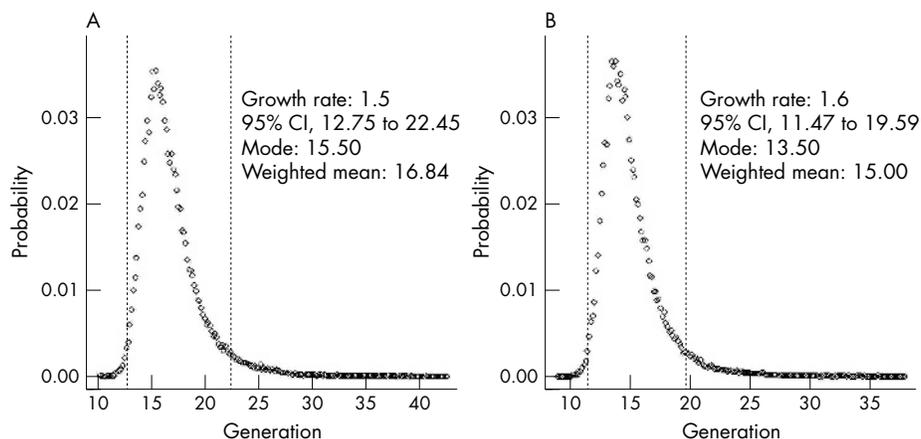


Figure 1 Plot of the posterior distribution of mutation age, estimated by the software DMLE⁸ (disease mapping using linkage disequilibrium) employing Markov Chain Monte Carlo estimation. The dotted lines show the 95% posterior confidence region for the number of generations.

Competing interests: none declared

Correspondence to: Dr Celia Greenwood, Genetics and Genomic Biology, Hospital for Sick Children, 555 University Ave, Toronto, Ontario, Canada M5G 1X8; celia.greenwood@utoronto.ca

Received 12 January 2005

Revised version received 12 February 2005

Accepted for publication 14 February 2005

REFERENCES

- 1 **National Cancer Institute.** *Canadian cancer statistics, 2004.* Toronto: National Cancer Institute of Canada, 2004.
- 2 **Kemp Z,** Thirlwell C, Sieber O, Silver A, Tomlinson I. An update on the genetics of colorectal cancer. *Hum Mol Genet*, 2004;**13** Spec No 2, R177–85.
- 3 **Foulkes WD,** Thiffault I, Gruber SB, Horwitz M, Hamel N, Lee C, Shia J, Markowitz A, Figar A, Friedman E, Farber D, Greenwood CM, Bonner JD, Nafa K, Walsh T, Marcus V, Tomsho L, Gebert J, Macrae FA, Gaff CL, Paillerets BB, Gregersen PK, Weitzel JN, Gordon PH, MacNamara E, King MC, Hampel H, de La Chapelle A, Boyd J, Offit K, Rennert G, Chong G, Ellis NA. The founder mutation MSH2*1906G→C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am J Hum Genet* 2002;**71**:1395–412.
- 4 **Ostler H.** A genetic profile of contemporary Jewish populations. *Nat Rev Genet* 2001;**2**:891–8.
- 5 **Risch N,** de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 1995;**9**:152–9.
- 6 **Labuda M,** Labuda D, Korab-Laskowska M, Cole DE, Zietkiewicz E, Weissenbach J, Popowska E, Pronicka E, Root AW, Glorieux FH. Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. *Am J Hum Genet* 1996;**59**:633–43.
- 7 **Rannala B,** Reeve JP. High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* 2001;**69**:159–78.
- 8 **Reeve JP,** Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* 2002;**18**:894–5.
- 9 **Kong A,** Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. *Nat Genet* 2002;**31**:241–7.
- 10 **Stephens M,** Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;**68**:978–89.
- 11 **Mitra N,** Ye TZ, Smith A, Chuai S, Kirchoff T, Peterlongo P, Nafa K, Phillips MS, Offit K, Ellis NA. Localization of cancer susceptibility genes by genome-wide single-nucleotide polymorphism linkage-disequilibrium mapping. *Cancer Res* 2004;**64**:8116–25.