Am. J. Hum. Genet. 59:1243–1251, 1996

Age and Origin of Two Common MLH1 Mutations Predisposing to Hereditary Colon Cancer

Anu-Liisa Moisio,[1] Pertti Sistonen,[2] Jean Weissenbach,[3] Albert de la Chapelle,[1] and Päivi Peltomäki[1]

[1]Department of Medical Genetics, Haartman Institute, University of Helsinki, and [2]Finnish Red Cross Blood Transfusion Center, Helsinki; and [3]Généthon and CNRS URA 1922, Evry

Summary

Two mutations in the DNA mismatch repair gene MLH1, referred to as *mutations 1* and *2*, are frequent among Finnish kindreds with hereditary nonpolyposis colorectal cancer (HNPCC). In order to assess the ages and origins of these mutations, we constructed a map of 15 microsatellite markers around MLH1 and used this information in haplotype analyses of 19 kindreds with mutation 1 and 6 kindreds with mutation 2. All kindreds with mutation 1 showed a single allele for the intragenic marker D3S1611 that was not observed on any unaffected chromosome. They also shared portions of a haplotype of 4–15 markers encompassing 2.0–19.0 cM around MLH1. All kindreds with mutation 2 shared another allele for D3S1611 and a conserved haplotype of 5–14 markers spanning 2.0–15.0 cM around MLH1. The degree of haplotype conservation was used to estimate the ages of these two mutations. While some recessive disease genes have been estimated to have existed and spread for as long as thousands of generations worldwide and hundreds of generations in the Finnish population, our analyses suggest that the spread of mutation 1 started 16–43 generations (400–1,075 years) ago and that of mutation 2 some 5–21 generations (125–525 years) ago. These datings are compatible with our genealogical results identifying a common ancestor born in the 16th and 18th century, respectively. Overall, our results indicate that all Finnish kindreds studied to date showing either mutation 1 or mutation 2 are due to single ancestral founding mutations relatively recent in origin in the population. Alternatively, the mutations arose elsewhere earlier and were introduced in Finland more recently.

Received July 8, 1996; accepted for publication September 19, 1996.

Address for correspondence and reprints: Dr. Päivi Peltomäki, Department of Medical Genetics, Haartman Institute, P. O. Box 21 (Haartmaninkatu 3), FIN-00014 University of Helsinki, Helsinki, Finland.

0002-9297/96/5906-0010$02.00

Introduction

Examples of disease-causing mutations that have existed for thousands of years include the CAG→TAG nonsense mutation at codon 39 in the beta-globin gene (Pirastu et al. 1987) and the ΔF508 mutation in the cystic fibrosis transmembrane conductance regulator gene (Kerem et al. 1989; Morral et al. 1994). The spread of these mutations to different human populations is likely to have been enhanced by a selective advantage conferred by heterozygosity for the mutation (Pirastu et al. 1987; Romeo et al. 1989). Although the a priori expectation in cancer might be that dominant predisposing mutations would be rapidly eliminated because of reduced fitness, this is often not the case because the age at onset may be past reproductive age. Thus, several recurrent mutations in the BRCA1 gene have already been documented and the ancestral origins have been dated 9–170 generations back in six (Neuhausen et al. 1996). The mechanisms by which such mutations spread and maintain their frequencies are not yet well understood. It is nevertheless of prime interest to determine, in each case, whether recurrent mutations are inherited or arise de novo; moreover, their occurrence and incidence in different populations may have important diagnostic implications.

Susceptibility to hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in one of four genes with DNA mismatch repair function (MSH2, MLH1, PMS1, and PMS2) (Leach et al. 1993; Bronner et al. 1994; Nicolaides et al. 1994; Papadopoulos et al. 1994). As a rule, the predisposing mutation varies from kindred to kindred, but a small number of recurrent mutations have been detected (de la Chapelle and Peltomäki 1995). One consists of a single-base change at the splice-donor site of MSH2 exon 5 and occurs in 12% of English kindreds (Froggatt et al. 1995). Two mutations in MLH1, referred to as *mutations 1* and *2*, are common in Finland and together account for ≤68% of kindreds fulfilling the Amsterdam criteria and 51% of all Finnish kindreds with verified or putative HNPCC (Nyström-Lahti et al. 1996). Mutation 1 consists of a large genomic deletion affecting exon 16 and flanking, introns while mutation 2 is a single-base change at the splice-acceptor site of exon 6.

In contrast to English kindreds with the MSH2 mutation, which may not share a common disease haplotype (Froggatt et al. 1995), we here show that all kindreds with mutation 1 and mutation 2, respectively, are likely to descend from a common ancestor, on the basis of the presence of large conserved disease haplotypes. According to our calculations, the spread of mutations 1 and 2, respectively, may have begun ~16–43 and 5–21 generations ago. These findings are discussed in view of the population history of the Finns.

Kindreds, Material, and Methods

Kindreds

Twenty-five Finnish HNPCC kindreds with either one of the two previously characterized mutations in MLH1 (Nyström-Lahti et al. 1995) were studied. Sixteen kindreds of 19 with mutation 1 (F1, F2, F3, F6, F9, F10, F11, F12, F19, F21, F30, F43, F58, F59, F66, and F77) met the international criteria for HNPCC (Vasen et al. 1991); three kindreds (F63, F69, and F82) did not. The mean number of affected individuals was 13 (range 3–33) and 5 (range 4–5) in the former and latter kindreds, respectively. The pedigrees of kindreds F2, F3, F6, F10, F11, and F59 have been published previously (Nyström-Lahti et al. 1994). All six kindreds with mutation 2 fulfilled the international criteria, with the average of six affected individuals per family (range 3–10). The mean age at onset was similar in kindreds with mutation 1 and 2: 44 years (range 37–55) and 45 years (range 43–50), respectively. The present kindreds are well representative of a larger series recently studied by Aarnio et al. (1995) for organ distribution and cumulative risk of cancer. This study has been approved by the Ethical Committee of the Department of Medical Genetics, University of Helsinki.

Genealogical Study

The ancestries of the families were traced for ≤ 18 generations by using data from the church parish registers; the records do not extend beyond that time. These nationwide registers provide details of births and deaths, marriages, and changes of residence. We have elsewhere presented some genealogical data of a proportion of the kindreds (Nyström-Lahti et al. 1995); we now extended these studies to further families and further past generations.

Haplotype Analysis

In seven kindreds (F2, F3, F6, F10, F11, and F59 with mutation 1 and F26 with mutation 2) the disease haplotype was constructed by genotyping the majority of all relevant (affected or unaffected) family members. Five to nine affected individuals and/or verified mutation carriers were included per family. In eight kindreds (F1, F12, F19, F21, and F66 with mutation 1 and F13, F27, and F55 with mutation 2) samples from a minimum of two affected individuals or verified mutation carriers and a variable number of noncarriers were available to determine reliably the disease haplotype. In the remaining 10 kindreds, a sample from only one affected individual was available. In these cases, phasing was not feasible, and the common haplotype shown in figure 1 represents a haplotype that is most likely based on the genealogical, mutational, and allelotyping data of these kindreds.

Fifteen highly polymorphic microsatellite markers from chromosome 3p (Dib et al. 1996) were analyzed. Genotyping data from eight CEPH families were available in the CEPH on-line database (http://www.cephb.fr/), and the relevant recombination fractions and their ± 1-LOD unit support values were retrieved from this source.

Because of insufficient resolution capacity of meiotic mapping, the Stanford G3 radiation hybrid panel developed by David Cox and purchased from Research Genetics was used in parallel to determine the distances between marker loci and their most likely order. These calculations were performed by using the RHMAP program package provided by Michael Boehnke (http://www.sph.umich.edu/group/statgen/software). First, maximum pairwise LOD scores and distance estimates were calculated by program RH2PT, and, in case of significant linkage (LOD score ≥ 3), ± 1-LOD-unit support intervals were calculated for the breakage frequencies between two markers. Second, the loci were ordered by multipoint analysis based on the maximum-likelihood method (Boehnke et al. 1991) using program RHMAXLIK. Since no hot spots for X-ray breakage are known to exist in the region around MLH1 on chromosome 3p, the obtained distances (in cR) were converted into kilobases by using the ratio of 1 cR/50 kb as suggested by Cox et al. (1990).

DNA samples were amplified by the PCR using primers specific for the above mentioned markers and the amplification products were separated through polyacrylamide gel electrophoresis as described in detail by Peltomäki et al. (1993). Alleles were numbered consecutively according to decreasing size.

Allelic Association

To assess the degree of linkage disequilibrium we determined the p_{excess} value: $p_{excess} = (p_D - p_N)/(1 - p_N)$, where p_D is the frequency of the associated allele on disease chromosomes and p_N is the frequency of the same allele on normal chromosomes. p_D was calculated from all verified disease chromosomes ($N = 58$ for mutation 1, and $N = 15$ for mutation 2). p_N was calculated from 50 normal chromosomes that either represented haplotypes not associated with HNPCC in the kindreds or originated from Finnish blood donors. Comparison of these two sources of normal chromosomes revealed

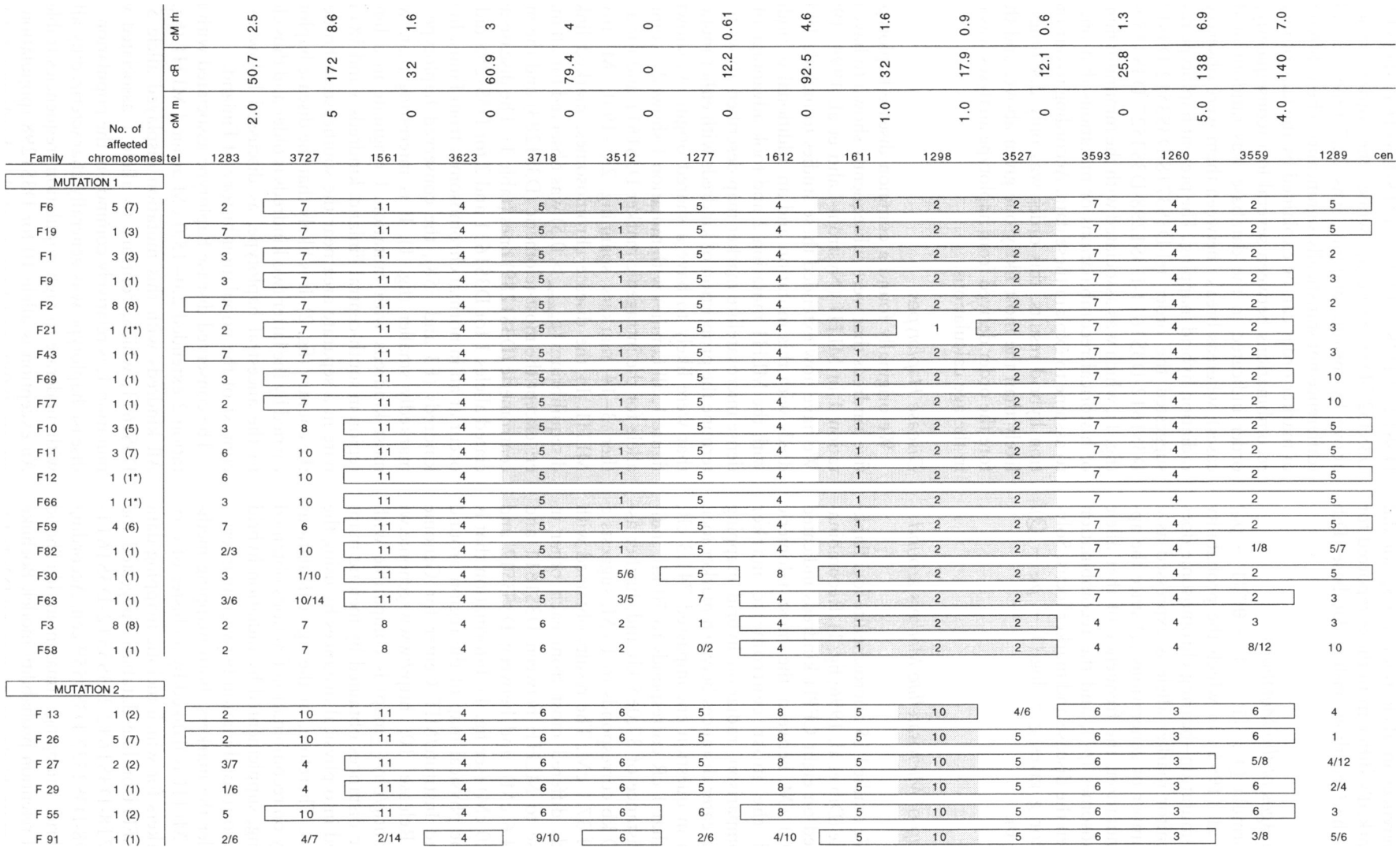

Figure 1 Disease haplotypes of kindreds with mutation 1 or mutation 2. Marker loci are shown from telomere (tel) to centromere (cen). The distances between markers are given (in cM) based on meiotic mapping (m) and (in centiRays [cR] and corresponding cM values [0.05 cM/1 cR]) based on radiation hybrid mapping (rh). The conserved parts of the haplotypes are boxed, and those consisting of alleles with low population frequencies are shaded. In the column showing the number of affected chromosomes, the numbers without parentheses indicate HNPCC chromosomes displaying exactly the haplotype shown, while those in parentheses indicate the total number of HNPCC chromosomes studied from each family. Asterisk (*) after 1 in this column indicates that, for a few markers, genotyping data are available from more than one affected individual.

no significant differences in allele frequencies, nor did the frequencies markedly differ from those reported for CEPH families in the Généthon on-line database (Dib et al. 1996).

Calculation of the Ages of the Mutations

We used the formula $Q = [1 - (1 - \theta)^g](1 - p_N)$ given by Risch et al. (1995), in which the probability (Q) that a disease-mutation–bearing chromosome does not carry a progenitor marker allele is expressed as a function of the number of generations (g) since the appearance of the mutation, the frequency of that allele on normal chromosomes (p_N), and the recombination fraction (θ) between the disease and marker locus. Solving the formula for g gives $g = \log[(1 - p_N - Q)/(1 - p_N)]/\log(1 - \theta)$.

Results

Constructing a Map of Microsatellite Markers around MLH1

Distance.—We studied 15 microsatellite loci flanking the MLH1 gene (Dib et al. 1996; fig. 1). Recombination mapping based on eight CEPH kindreds indicated a distance of 19.0 cM between the most telomeric (D3S1283) and the most centromeric marker (D3S1289). In comparison, radiation hybrid mapping generally yielded greater distances between marker loci. The whole region in question encompassed 741.3 cR, which, assuming that 1 cR corresponds to 50 kb, indicates a physical distance of 37,065 kb and, further assuming that 1,000 kb corresponds to 1 cM, suggests a genetic distance of 37.1 cM. The results obtained with these two methods differed most from each other in two intervals close to MLH1: between D3S1611 and D3S1612 (0 vs. 4.6 cM), and between D3S3512 and D3S3623 (0 vs. 7.0 cM) (see fig. 1). Information that is available about the physical map of the region around MLH1 (Whitehead Institute/MIT Center for Genome Research, Data Release 10; http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map) is compatible with the present genetic distances obtained by meiotic mapping and converted into physical distances by using the 1,000 kb/cM ratio. Therefore, in the age calculations we relied primarily on recombination fractions obtained by meiotic mapping, supplemented by radiation hybrid data in cases of zero recombination between markers.

Order.—To order the markers, both mapping methods were applied. MLH1 is flanked by a cluster of very closely linked markers for which meiotic mapping data available in the CEPH on-line database suggested the order tel-D3S3718-D3S1612-D3S3512-D3S1611-D3S1277-D3S1298-D3S3527-D3S3559-cen. According to radiation hybrid mapping (maximum-likelihood method with equal retention probability model; Boehnke et al. 1991), the most likely order was tel-D3S1612-D3S1277-D3S3512-D3S3718-D3S1611-D3S1298-D3S3527-D3S3559-cen, while the order indicated above by meiotic mapping was 22.7 times less likely (difference not statistically significant). These differences in marker order can be explained by the lack of meiotic recombinations between several loci (consequently, the order indicated in the database was random in these cases), whereas breaks between them were observed by radiation hybrid mapping. The present haplotype results suggested the order tel-D3S3718-D3S3512-D3S1277-D3S1612-D3S1611-D3S1298-D3S3527-D3S3559-cen (fig. 1), which was associated with the minimum number of historical recombinations or mutations that one has to assume to explain these data. According to our radiation hybrid results, this order was only 2.4 times less likely than the most likely one given above, and therefore the order derived from haplotype analyses was used in the age calculations.

Disease Haplotypes

We originally found a common disease haplotype in five kindreds that were subsequently shown to have mutation 1 in MLH1 (Nyström-Lahti et al. 1994, 1995). We have now extended these studies to include 19 kindreds with this mutation and an additional 6 kindreds with the MLH1 mutation 2 and took advantage of the dense microsatellite marker map described above to determine the proportion of kindreds with either mutation that were likely to have a shared origin. As shown in figure 1, all families with mutation 1 showed a common allele for the intragenic marker D3S1611 and for a further 3–14 markers spanning 2.0–19.0 cM around MLH1. In the disease chromosomes, marked linkage disequilibrium ($p_{excess} > 0.5$) was observed for all markers except the most-telomeric (D3S1283) and the most-centromeric (D3S1289) ones (table 1). The disease-associated alleles 1 for D3S1611 and 2 for D3S1298 did not occur at all in normal chromosomes from Finns. In two kindreds (F3 and F58), the conserved haplotype was markedly smaller (fig. 1). It is interesting that, while most kindreds with mutation 1 originate in a limited region in south-central Finland, kindreds 3 and 58 originate in a separate area near the southeastern border of the country (fig. 2). It is likely that the disease haplotype carried by the latter two kindreds is older and thus closer to the ancestral haplotype, as discussed below in the context of the population history of Finland.

The conserved disease haplotype associated with mutation 2 extended 2.0–15.0 cM around MLH1 (fig. 1). All kindreds with this mutation displayed allele 5 for D3S1611, which, contrary to allele 1 associated with mutation 1, is relatively common in the population. The disease haplotype was generally characterized by alleles with moderate to high population frequencies (table 1). An exception is allele 10 for D3S1298 (population frequency 0.08), and because of this, haplotype 5-10 for

Table 1

Linkage Disequilibrium between MLH1 and Flanking Markers Assessed by Allelic Excess, i.e., p_{excess}

	Mutation 1					Mutation 2				
Marker	Allele	Size (bp)	p_N	p_D	p_{excess}	Allele	Size (bp)	p_N	p_D	p_{excess}
D3S1283	7	148	.080	.276	.213	2	158	.440	.500	.107
D3S3727	7	127	.140	.586	.519	10	121	.280	.733	.630
D3S1561	11	226	.440	.828	.692	11	226	.440	.933	.881
D3S3623	4	217	.380	1.000	1.000	4	217	.380	1.000	1.000
D3S3718	5	160	.040	.828	.820	6	158	.440	.933	.881
D3S3512	1	143	.060	.810	.798	6	133	.240	1.000	1.000
D3S1277	5	268	.200	.821	.777	5	268	.200	.800	.750
D3S1612	4	218	.360	.982	.973	8	210	.420	.933	.885
D3S1611	1	270	.000	1.000	1.000	5	262	.400	1.000	1.000
D3S1298	2	212	.000	.930	.930	10	198	.080	1.000	1.000
D3S3527	2	117	.020	.947	.946	5	111	.160	.867	.841
D3S3593	7	213	.120	.759	.726	6	215	.340	1.000	1.000
D3S1260	4	264	.160	.948	.938	3	266	.540	1.000	1.000
D3S3559	2	187	.120	.737	.701	6	187	.100	.800	.778
D3S1289	5	215	.120	.345	.256	ND	...	...	...	...

Note.—D3S1611 is located within MLH1. ND = not determined.

D3S1611-D3S1298, present in all disease chromosomes from kindreds with mutation 2, occurred in only 2 (4%) of 50 normal chromosomes. Despite high p_N values, linkage disequilibrium was observable over a large genetic distance (table 1). As shown in figure 2, the geographic region of origin of kindreds with mutation 2 is different from that of kindreds with mutation 1.

Calculating the Ages of the Mutations

The frequencies of various disease haplotypes among the kindreds were used to estimate the time since the spread of these two mutations began. As detailed in Kindreds, Material, and Methods, we applied the formula that Risch et al. (1995) previously used to date a mutation predisposing to another autosomal dominant disorder, idiopathic torsion dystonia, in Ashkenazi Jews. The results of age calculations are shown in table 2. For these calculations, we used p_N values given in table 1 and θ values obtained from genetic distances given in figure 1, derived from either meiotic recombination or radiation hybrid mapping. Age calculations were carried out in two ways. First, we used only one (the oldest and/or the most-common) disease haplotype (fig. 1) to represent each kindred ("representatives"; $N = 19 + 6$). Second, we included all individuals, either affected by cancer or shown to be mutation carriers by our gene analyses, who shared the same core haplotype ("all"; $N = 58 + 15$). The latter approach takes into account known recombination events within families.

In kindreds with mutation 1, the use of only one representative haplotype per family (and θ values from meiotic mapping, if possible) resulted in estimates varying from 15.4 (6.2–78.8) to 27.0 (11.7–109.8) generations since the beginning of mutation spread. Taking all available samples into account, the estimated numbers of generations were somewhat higher (16.4–42.6). On the assumption of an average generation length of 25 years, the spread of mutation 1 may thus have started ~400–1,075 years ago. In these calculations, our interpretations of the disease haplotypes shown in figure 1 were conservative, to minimize the effect of possible historical recombinations that could not be confirmed, mainly when samples from no more than one affected individual were available. For example, we assumed that, in kindred 21, allele 1 at D3S1298 represents a mutation, most likely involving a gain of one repeat unit to allele 2 as a result of polymerase slippage at DNA replication. We further assumed that, in kindreds F30 and F63, the conserved haplotype extends from D3S1289 and D3S3559, respectively, to D3S1561, since both kindreds displayed the rare disease-associated allele 5 at D3S3718.

Somewhat lower age estimates were obtained for mutation 2 (table 2). It was more difficult to date this mutation accurately, because of the relatively small number of disease haplotypes. When we excluded the lowest estimate (7.0/2.6), which was apparently biased by the surprisingly high recombination fraction (4.6) obtained for the D3S1611-D3S1612 interval by radiation hybrid mapping (see above), the estimates based on "representatives" ranged from 9.0 to 24.7 generations, while those based on "all" samples varied between 4.9 and 20.9, with wide support intervals. On the basis of the latter values, the spread of mutation 2 may therefore have begun some 125–525 years ago.

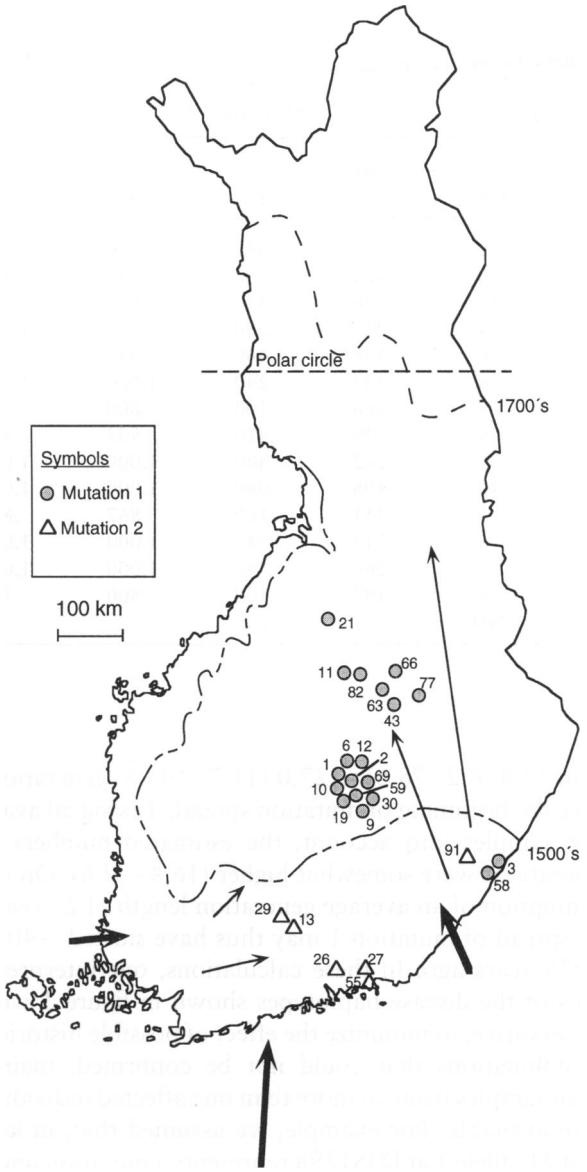

Figure 2 The ancestral geographic origin of kindreds with mutation 1 or mutation 2. Arrows indicate movements resulting in the initial founding (in bold) and stepwise inhabitation of different regions of Finland (adapted from Nevanlinna [1972]; Halila [1985]; Suvanto [1985]). Northern borders of settlement in the beginning of the 1500s and 1700s are shown separately by dashed lines.

The age estimates given above were found to be well in agreement with our genealogical findings, which pinpointed a common ancestor born in 1505 for kindreds 1, 6, 9, 10, and 30, with mutation 1 (Nyström-Lahti et al. 1994), while kindreds 27 and 55, with mutation 2, had a common ancestor born in the early 18th century (present study).

Discussion

Traditionally, analysis of disease-associated haplotypes and linkage disequilibrium has been successfully used for high-resolution mapping of as yet unidentified disease genes in suitable populations (Hästbacka et al. 1992; Ozelius et al. 1992; Hellsten et al. 1993; Lehesjoki et al. 1993; Nyström-Lahti et al. 1994; Höglund et al. 1995; Varilo et al. 1996). As soon as accurate information is available about genetic and/or physical distances between the disease locus and flanking marker loci, a reverse approach can be used to analyze the age and origin of recurrent disease-associated mutations if the population growth can be modeled, as exemplified by the present study and previous investigations by others (Risch et al. 1995; Neuhausen et al. 1996). While the proportion of HNPCC of the total colorectal cancer burden in Finland is not estimated to be any higher than in other industrialized countries (Aaltonen et al. 1994; Mecklin et al. 1995), a remarkable feature of Finnish HNPCC kindreds is that many of them are due to two mutations in MLH1. On the basis of conserved disease haplotypes, we show here that all Finnish kindreds with mutation 1 are likely to descend from a common ancestor; the same is true kindreds with mutation 2. Furthermore, our results suggest that these mutations are relatively recent in origin, which is supported by the observed extensive haplotype conservation around MLH1 in disease chromosomes.

For age calculations, we used a formula that includes three parameters (Q, p_N, and θ; see Kindreds, Material, and Methods) that are potential sources of errors because of possible sampling bias (Q and p_N) or uncertainty in the estimates of θ (Risch et al. 1995). To minimize the sampling variance, we applied two different methods to obtain values for Q (see Results) and p_N (see Kindreds, Material, and Methods). To reduce the effect of inaccurate estimates of θ, we determined the distances between adjacent marker loci in two ways, by meiotic recombination and radiation hybrid mapping, and calculated the 1-LOD-unit support intervals for each value of θ, which were then used to determine the support intervals for the numbers of generations (g) since the appearance of the mutation. Contrary to Neuhausen et al. (1996), we intentionally chose a formula that does not depend on mutation rates of microsatellite markers, since it would have been difficult to estimate these rates in individuals with DNA mismatch-repair gene mutations. Finally, in addition to the formula by Risch et al. (1995) we also tested the formula by Hästbacka et al. (1992), which was found to yield comparable age estimates.

As far as we know, neither mutation 1 nor mutation 2 has been found in non-Finnish kindreds. However, the fact that they have not yet been found elsewhere does not mean that they do not exist. A look at the map (fig. 2) and the apparent high likelihood that the disease haplotype of families 3 and 58 is older than that of the remaining families with mutation 1 make it well worth considering whether this mutation might have been in-

Table 2

Estimates of the Ages of Mutations 1 and 2

Loci and Group	Proportion Recombinant	Generations from[a]: Radiation Hybrid Mapping	Generations from[a]: Meiotic Mapping	θ_{rh}	θ_m
			Mutation 1		
D3S1277–D3S1612:					
Representatives	2/19 = .105	23.1 (10.7–100.7)	No value[b]	.006 (.001–.013)	.000 (.000–.020)
All	9/58 = .155	35.3 (16.4–153.9)	No value[b]		
D3S3727–D3S1561:					
Representatives	8/17 = .471	8.8	15.4 (6.2–78.8)	.086	.050 (.010–.120)
All	24/49 = .490	9.4	16.4 (6.6–83.9)		
D3S3593–D3S3527:					
Representatives	2/19 = .105	21.1 (10.6–159.2)	No value[b]	.006 (.001–.012)	.000 (.000–.030)
All	11/55 = .200	42.6 (21.4–322.2)	No value[b]		
D3S1289–D3S3559:					
Representatives	10/17 = .588	15.1	27.0 (11.7–109.8)	.070	.040 (.010–.090)
All	22/41 = .537	13.1	18.3 (8.1–93.6)		
			Mutation 2		
D3S1612–D3S1611:					
Representatives	1/6 = .167	7.0	No value[b]	.006	.000 (.000–.020)
All	1/15 = .067	2.6	No value[b]		
D3S3727–D3S1561:					
Representatives	2/5 = .4	9.0	15.8 (6.3–80.7)	.086	.050 (.010–.120)
All	3/14 = .214	3.9	6.8 (2.7–53.2)		
D3S1283–D3S3727:					
Representatives	1/3 = .333	35.2 (19.2–84.1)	24.7 (10.8–no value)[b]	.025 (.0107–.0461)	.020 (.000–.080)
All	3/10 = .3	29.8 (16.3–71.3)	20.9 (89.2–no value)[b]		
D3S3559–D3S1260:					
Representatives	2/6 = .333	6.5	9.0 (4.0–46.0)	.069	.050 (.010–.110)
All	3/15 = .2	3.5	4.9 (2.1–25.0)		

Note.—rh = radiation hybrid mapping; m = meiotic mapping. See text for details.

[a] Support intervals based on 1-LOD-unit support intervals for the recombination fractions are given in parentheses.

[b] Using the formula by Risch et al. (1995), the value of g cannot be determined if θ is 0.

troduced into Finland from the southeast. There are at least two obvious reasons for the absence of reports of the mutation outside Finland: First, mutation data available so far almost exclusively come from western populations, and, second, this particular mutation is prone to elude detection by commonly used screening techniques (Nyström-Lahti et al. 1995). On the basis of our genealogical findings, it is possible that mutation 2, too, was introduced into Finland from the East, as discussed below.

We propose that genetic drift in an isolated population is the most-likely mechanism by which these two mutations were enriched in the Finnish population, in analogy with the enrichment of many other (mainly recessive) disease genes in Finns (Vogel and Motulsky 1986; de la Chapelle 1993). A similar mechanism may have resulted in the high frequencies of particular disease-causing mutations among Ashkenazi Jews (Motulsky 1995; Risch et al. 1995). While a few recessive disease-causing mutations occurring throughout Finland have probably existed and expanded for $\geq$100 generations in this population (Hästbacka et al. 1992), other recessive mutations that occur in more restricted areas range between 15 and 30 generations (Höglund et al. 1995; Varilo et al. 1996). Archaeological and genetic evidence suggests that Finland was populated beginning some 9,000 years ago by settlers coming both from the South (southwest) and from the East (southeast) (Nevanlinna 1972; Meinander 1984; Sajantila et al. 1995) The main expansion resulting in today's population of 5 million Finns is thought to have started some 100 generations (2,000–2,500 years) ago from a founding population of very limited size (~2,500) (Jutikkala 1984; de la Chapelle 1993). The small number of ancestors allows strong founder effects; moreover, genetic drift as a result of population bottlenecks and random events can lead to profound changes in gene frequencies.

Until the 16th century, the settlement was mainly re-

stricted to the coastal and southern parts of the country, leaving the eastern and northern parts relatively unpopulated (fig. 2). These areas served as hunting and fishing grounds for people living in southern Finland. In the eastern part of the country the land was cleared for cultivation by burning, and, since it was possible to get a good harvest only up to three times by this means, these people were gradually forced to move northward. A more permanent settlement was established in the central and northern parts of the country, as late as the early 16th century, followed by a regional population expansion starting in the 17th century (Kiuasmaa 1985). The core region in which most kindreds with mutation 1 originate (fig. 2) represents a single large historical parish that was founded in the 1550s, and the estimated beginning of the spread of mutation 1 coincides with the above described internal northward migration and the subsequent population expansion in these areas. The spread of mutation 1 may thus be characterized by at least two bottlenecks, one at the earlier introduction in the southeast (families 3 and 58) and one at the 1550s founding of the settlement 300 km to the northwest.

Less accurate information is available about the population history of the southeastern part of the country, where the ancestors of kindreds with mutation 2 lived. This is in part due to numerous wars against Russia in the 1600–1700s that resulted in several changes in the eastern border of Finland and also led to the loss of many genealogically important records and documents. We were able to trace a common ancestor for two kindreds (27 and 55), born in the early 18th century, to Russian Karelia (East of where the family members live today), suggesting a possible eastern origin of mutation 2. As more and more HNPCC kindreds are subjected to mutation studies worldwide, it will be interesting to see whether mutations 1 and 2 will be detected in kindreds not known to be of Finnish origin and, if so, whether haplotype analyses in those cases will be compatible with the present conclusions about the age and origin of these mutations.

Acknowledgments

We thank Dr. Juha Kere for helpful discussions; Sari Jokela-Hartikainen for genealogical studies; and Siv Lindroos, Sinikka Lindh, Kirsi Hopponen, Tuula Lehtinen, and Kaija Collin for sample collection and preparation. This work was financially supported by the Academy of Finland, the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the Nordic Cancer Union, and the Federation of the Finnish Insurance Companies. Part of the investigation was conducted at the Folkhälsan Institute of Genetics.

References

Aaltonen LA, Sankila R, Mecklin J-P, Järvinen H, Pukkala E, Peltomäki P, de la Chapelle A (1994) A novel approach to estimate the proportion of hereditary non-polyposis colorectal cancer of total colorectal cancer burden. Cancer Detect Prev 18:57–63

Aarnio M, Mecklin J-P, Aaltonen LA, Nyström-Lahti M, Järvinen H (1995) Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 64:430–433

Boehnke M, Lange K, Cox DR (1991) Statistical methods for multipoint radiation hybrid mapping. Am J Hum Genet 49: 1174–1188

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, et al (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 368:258–261

Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990) Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245–250

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

de la Chapelle A, Peltomäki P (1995) Genetics of hereditary colon cancer. Annu Rev Genet 29:329–349

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Froggatt NJ, Joyce JA, Davies R, Evans DGR, Ponder BAJ, Barton DE, Maher ER (1995) A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome. Lancet 345:727

Halila A (1985) Karoliinisen ajan väestö ja talous. In: Avikainen P, Laaksonen E, Pärssinen E (eds) Suomen historia, part III. Weilin+Göös, Espoo, pp 335–337

Hellsten E, Vesa J, Speer M, Mäkelä TP, Järvelä I, Alitalo K, Ott J, et al (1993) Refined assignment of the infantile neuronal lipofuscinosis (INCL, CLN1) locus at 1p32: incorporation of linkage disequilibrium in multipoint analysis. Genomics 16:720–725

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204–211

Höglund P, Sistonen P, Norio R, Holmberg C, Dimberg A, Gustavson K-H, de la Chapelle A, et al (1995) Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. Am J Hum Genet 57:95–102

Jutikkala E (1984) A summary of the discussion on "The roots of the Finnish people." In: Åstrom S-E (ed) Suomen väestön esihistorialliset juuret. Tvärminnen symposiumi 17.-19.1.1980. Societas Scientiarum Fennica, Helsinki, pp 365–376

Kerem B, Rommens JM, Buchanan JA, Markiewitcz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245: 1073–1080

Kiuasmaa K (1985) Erämaat ja kaupungit. In: Avikainen P, Laaksonen E, Pärssinen E (eds) Suomen historia, part II. Weilin+Göös, Espoo, pp 335–337

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Par-

sons R, Peltomäki P, et al (1993) Mutations of a MutS homolog in hereditary non-polyposis colorectal cancer. Cell 75:1215–1225

Lehesjoki AE, Koskiniemi M, Norio R, Tirrito S, Sistonen P, Lander E, de la Chapelle A (1993) Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. Hum Mol Genet 2:1229–1234

Mecklin J-P, Järvinen HJ, Hakkiluoto A, Hallikas H, Hiltunen K-M, Härkönen N, Kellokumpu I, et al (1995) Frequency of hereditary nonpolyposis colorectal cancer. A prospective multicenter study in Finland. Dis Colon Rectum 38:588–593

Meinander CF (1984) The Stone Age ancestors. In: Åstrom S-E (ed) Suomen väestön esihistorialliset juuret. Tvärminnen symposiumi 17.-19.1.1980. Societas Scientiarum Fennica, Helsinki, pp 21–48

Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Giménez J, Reis A, et al (1994) The origin of the major cystic fibrosis mutation (ΔF508) in European populations. Nat Genet 7:169–175

Motulsky AG (1995) Jewish diseases and origins. Nat Genet 9:99–101

Neuhausen SL, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlison G, et al (1996) Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. Am J Hum Genet 58: 271–280

Nevanlinna HR (1972) The Finnish population structure: a genetic and genealogic study. Hereditas 71:195–236

Nicolaides NC, Papadopoulos N, Liu B, Wei Y-F, Carter KC, Ruben SM, Rosen CA, et al (1994) Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 371:75–80

Nyström-Lahti M, Kristo P, Nicolaides NC, Chang S-Y, Aaltonen LA, Moisio A-L, Järvinen HJ, et al (1995) Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nat Med 1:1203–1206

Nyström-Lahti M, Sistonen P, Mecklin J-P, Pylkkänen L, Aaltonen LA, Järvinen H, Weissenbach J, et al (1994) Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proc Natl Acad Sci USA 91:6054–6058

Nyström-Lahti M, Wu Y, Moisio A-L, Hofstra RMW, Osinga J, Mecklin J-P, Järvinen HJ, et al (1996) DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Hum Mol Genet 5:763–769

Ozelius LJ, Kramer PL, de Leon D, Risch N, Bressman SB, Schuback DE, Brin MF, et al (1992) Strong allelic association between the torsion dystonia gene (*DYT1*) and loci on chromosome 9q34 in Ashkenazi Jews. Am J Hum Genet 50: 619–628

Papadopoulos N, Nicolaides NC, Wei Y-F, Ruben SM, Carter KC, Rosen CA, Haseltine WA, et al (1994) Mutation of a mutL homolog is associated with hereditary colon cancer. Science 263:1825–1829

Peltomäki P, Aaltonen LA, Sistonen P, Pylkkänen L, Mecklin J-P, Järvinen H, Green JS, et al (1993) Genetic mapping of a locus predisposing to human colorectal cancer. Science 260:810–812

Pirastu M, Galanello R, Doherty MA, Tuveri T, Cao A, Kan YW (1987) The same β-globin gene mutation is present on nine different β-thalassemia chromosomes in a Sardinian population. Proc Natl Acad Sci 84:2882–2885

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, et al (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152–159

Romeo G, Devoto M, Galietta JV (1989) Why is the cystic fibrosis gene so frequent? Hum Genet 84:1–5

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus M-L, Aula P, et al (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5:42–52

Suvanto S (1985) Keskiajan summa. In: Avikainen P, Laaksonen E, Pärssinen E (eds) Suomen historia, part II. Weilin+Göös, Espoo, pp 217–223

Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Järvelä I (1996) The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet 58:506–512

Vasen HF, Mecklin J-P, Meera Khan P, Lynch HT (1991) The international collaborative group on hereditary nonpolyposis colorectal cancer (ICG-HNPCC). Dis Colon Rectum 34:424–25

Vogel F, Motulsky AG (1986) Rare flora in rare soil: hereditary diseases in Finland. In: Vogel F, Motulsky AG (eds) Human genetics, 2d ed. Springer, Berlin, pp 509–511