Am. J. Hum. Genet. 59:633–643, 1996

Linkage Disequilibrium Analysis in Young Populations: Pseudo–Vitamin D–Deficiency Rickets and the Founder Effect in French Canadians

Małgorzata Labuda,[1,2] Damian Labuda,[4,5] Maria Korab-Laskowska,[4] David E. C. Cole,[6] Ewa Zietkiewicz,[4] Jean Weissenbach,[7] Ewa Popowska,[8] Ewa Pronicka,[8] Allen W. Root,[9] and Francis H. Glorieux[1,2]

[1]Genetics Unit, Shriners Hospital for Crippled Children, [2]Departments of Surgery and Pediatrics and [3]Department of Human Genetics, McGill University, [4]Centre de Cancérologie Charles Bruneau, Centre de Recherche, Hôpital Sainte-Justine, and [5]Département de Pédiatrie, Université de Montréal, Montreal; [6]Department of Clinical Biochemistry, Banting Institute, Toronto; [7]Généthon, CNRS URA 1922, Evry, France; [8]Centrum Zdrowia Dziecka, Warsaw; and [9]University of South Florida, Tampa

Summary

Pseudo–vitamin D–deficiency rickets (PDDR) was mapped close to D12S90 and between proximal D12S312 and distal (D12S305,D12S104) microsatellites that were subsequently found on a single YAC clone. Analysis of a complex haplotype in linkage disequilibrium (LD) with the disease discriminated among distinct founder effects in French Canadian populations in Acadia and in Charlevoix-Saguenay-Lac-Saint-Jean (Ch-SLSJ), as well as an earlier one in precolonial Europe. A simple demographic model suggested the historical age of the founder effect in Ch-SLSJ to be ~12 generations. The corresponding LD data are consistent with this figure when they are analyzed within the framework of Luria-Delbrück model, which takes into account the population growth. Population sampling due to a limited number of first settlers and the rapid demographic expansion appear to have played a major role in the founding of PDDR in Ch-SLSJ and, presumably, other genetic disorders endemic to French Canada. Similarly, the founder effect in Ashkenazim, coinciding with their early settlement in medieval Poland and subsequent expansion eastward, could explain the origin of frequent genetic diseases in this population.

Introduction

Pseudo–vitamin D–deficiency rickets (PDDR), also referred to as "vitamin D dependency type I" (MIM 264700), is an autosomal recessive disorder characterized by hypocalcemia, secondary hyperparathyroidism, and early onset of rickets (Prader et al. 1961). Diagnosis is confirmed by determination of very low or undetectable levels of 1,25-dihydroxyvitamin D, the active form of vitamin D (Fraser et al. 1973; Delvin et al. 1981). PDDR results from renal 25(OH)-vitamin D 1α-hydroxylase deficiency, but the underlying molecular defect remains unknown. It could be due to a mutation in cytochrome P450cc1α, the moiety conferring specificity to this three-subunit enzyme. However, a defect in a gene regulating its expression is another possibility. Cytochrome P450cc24 was excluded as a candidate gene for PDDR by its localization to 20q13.3 (Labuda et al. 1993), whereas the disease has been mapped, by linkage analysis, to 12q13-14, close to the collagen (COL2A1) locus and to a haplotype consisting of three anonymous markers, hereafter referred to as "3MH" (Labuda et al. 1990).

In order to refine PDDR localization for positional cloning, we pursued linkage analysis by family studies and by the linkage disequilibrium (LD) approach. LD, which has played an important role in the fine mapping of cystic fibrosis (Kerem et al. 1989) and in localization of the vitamin D receptor adjacent to the COL2A1 gene (Labuda et al. 1992), takes advantage of the fact that all affected chromosomes originating from a common ancestor share the same alleles for markers in the vicinity of the disease locus. This set of neighboring alleles forms the ancestral or "founder" haplotype, which is expected to decay over the generations, with the alleles of the outer markers being exchanged more readily by recombination. The gain in mapping resolution by LD analysis is due to the increased number of meioses, *M*, analyzed, when the information from all generations following the founder event is used (Hästbacka et al. 1992; Lehesjoki et al. 1993; Ellis et al. 1994; Sulisalo et al. 1994; Mitchison et al. 1995; Levy et al. 1996; Varilo et al. 1996). Alternatively, if genetic distances are known, LD analysis allows one to estimate the age of founder effect(s) (Bodmer 1972; Serre et al. 1990; Risch et al. 1995*b*).

Hästbacka et al. (1992) recognized that the quantitative results of LD analysis are altered in the population

Received January 9, 1996; accepted for publication June 10, 1996.
Address for correspondence and reprints: Dr. Malgorzata Labuda, Genetics Unit, Shriners Hospital for Crippled Children, 1529 Cedar Avenue, Montreal, Quebec H3G 1A6, Canada.
© 1996 by The American Society of Human Genetics. All rights reserved.
0002-9297/96/5903-0019$02.00

undergoing demographic expansion, which is typically the case when the founder effect is observed. To calculate the likely genetic distances from their LD data in the Finnish population, they applied the model of Luria and Delbrück (1943), originally developed for the analysis of mutations in growing bacterial cultures. Their work, followed by that of others (Kaplan et al. 1995; Terwilliger 1995), also pointed to the importance of demographic effects in the founding of genetic diseases in expanding populations.

As in the Finns (de la Chapelle 1993) and Ashkenazim (Motulsky 1995), so in French Canadians, where more than a dozen otherwise rare genetic diseases are common, the unique genetic profile can be related to their demographic history. French colonization of eastern Canada began in the early 17th century. A region of maritime Canada known as Acadia was first colonized in the early 1600s by direct immigration from France and by subsequent rapid growth (Arsenault 1978). The first settlement, which gave rise to the Acadian population, was at Port Royal (now Annapolis, Nova Scotia), and in 1671 it had 373 inhabitants (Arsenault 1978). The population grew to 2,000 in 1713 and further expanded to 18,000 inhabitants in 1750. With the British conquest in 1755, >10,000 Acadians were deported along the eastern coast of the United States as far as Louisiana, with a resulting steep decline of the Acadian population in Canada (8,000 in 1800). A large part of this population emigrated to neighboring regions of New Brunswick, Cape Breton, and Prince Edward Island, and over time it grew to 139,000 in 1901 and to 286,000 in 1951. Now Canadians claiming Acadian descendants number 357,000 (Canadian Encyclopedia 1984).

Charlevoix, a geographically isolated area on the northern shore of the Saint Lawrence river, was colonized later, beginning in 1675, by a small number of founders (599 total; for details, see Subjects and Methods and fig. 1) arriving from the region of Quebec City over a period of seven generations (Jetté et al. 1991). Beginning in 1838, the familial migration from Charlevoix also populated the shores of the Saguenay river, the Lac-Saint-Jean area, and the Haute-Côte-Nord (Gauvreau et al. 1991). This population, referred to as that of "Ch-SLSJ," grew from 18,000 in 1852 to >65,000 in 1911 and to ~310,000 today.

The founder effect of PDDR in French Canadians has been indicated by genetic data (Labuda et al. 1990), consistent with the carrier frequency of 1/26 in the Ch-SLSJ region (De Braekeleer 1991), ~1/60 in the Acadian population (D.E.C.C., unpublished observation), and <1/500 elsewhere (as estimated in the Polish population). Other genetic disorders that are rare or unknown elsewhere also have been found at elevated frequency in the Ch-SLSJ region. They include spastic ataxia, with a carrier frequency of 1/21; tyrosinemia I, 1/22; polyneuropathy, 1/25; cytochrome C oxidase deficiency, 1/28; sarcosinemia, 1/29; histidinemia, 1/32; cystinosis, 1/39; mucolipidosis, 1/39; lipoprotein lipase deficiency, 1/43; and pyruvate kinase deficiency, 1/64 (De Braekeleer 1991; Gahl et al. 1995; Kornfeld and Sly 1995; Levy et al. 1995; Mitchell et al. 1995; Scott 1995; Tanaka and Pagalia 1995).

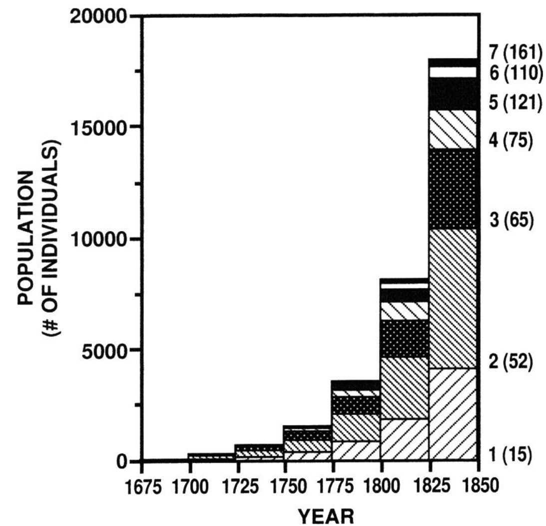

Figure 1 Demographic expansion in Ch-SLSJ between 1675 and 1850. The genetic contribution of each of seven immigration waves (designated "1"–"7" on the right) to the total population in a given time interval is highlighted by a different pattern in the histogram. The number of founders (Jetté et al. 1991) within each wave of immigrants is given in parentheses. The value $d = .8$ was obtained by simple iteration from the data (see Subjects and Methods), from knowing the total number of inhabitants in 1852 (Jetté et al. 1991).

In this study we provide the assignment of PDDR between markers located on a single YAC clone. Analysis of a complex haplotype has been used to detect LD and to discriminate among founder effects in Acadia and Ch-SLSJ, where demographic effects can account for the founding of rare mutations including PDDR.

Subjects and Methods

Families

Thirty-two nuclear PDDR families, which included 49 patients, were studied. Ten of these families have been described elsewhere (Labuda et al. 1990). All Quebec families (n = 20) originated from (on the basis of parental or grandparental birthplaces) Ch-SLSJ. Acadian families were from Prince Edward Island (n = 3) and New Brunswick (n = 2). One family was of mixed Quebec–New Brunswick origin, and one from Prince Edward Island was of non-Acadian ancestry. Three families were from Poland; and two were from Florida, one each of European and Filipino origin.

Genetic Typing

The DNA marker loci used are listed in table 1; the typing of 3MH and COL2A1 has been described else-

Table 1

Two-Point Linkage Analysis Between PDDR and 12q13-14 Markers

Locus	θ[a]	Maximum *Z*
COL2A1	.120	2.99
3MH	.086 (.023–.192)	4.66
D12S361	.072 (.020–.195)	3.35
D12S96	.030 (.010–.140)	4.05
D12S325	.051 (.011–.150)	4.43
D12S103	.138	1.48
D12S359	.097 (.035–.220)	3.34
D12S312	.025 (.001–.119)	5.27
D12S90	.000 (.042)	7.49
D12S305	.027 (.001–.128)	4.65
D12S104	.026 (.001–.140)	3.35
D12S355	.036 (.007–.109)	6.99
D12S83	.056 (.015–.144)	5.37

[a] Data in parentheses are bounds obtained as values at the maximum *Z* minus the LOD unit.

where (Labuda et al. 1990, 1992). All analyzed microsatellite markers were from Gyapay et al. (1994). For the marker D12S312 a primer pair different from the published one has been used: forward, GGTAATTATGATTGTTGTTAGAG; and reverse, GCAGGCATTTGTAAGCATATA. All PCR reactions were carried out in a total volume of 20 μl in 20 mM Tris-HCl pH 8.4, 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 μM each primer (one labeled with [γ]-^{32}P-ATP, used at 0.5×10^6 cpm/reaction), 40 ng genomic or YAC DNA, and 0.5–1 U *Taq* polymerase (Perkin Elmer). The incubation cycle, 94°C/55°C/72°C for 45 s each, was repeated 30 times. Alleles were numbered as follows: the longest variant identified was assigned number "1," and all the shorter ones were numbered by adding the number of CA dinucleotides by which they differed from allele 1. Haplotypes were constructed by using the most parsimonious phase, by tracing segregation of alleles in families, and by using the published map order (fig. 2*b*) (Gyapay et al. 1994), subsequently refined in light of our results (fig. 2*c*).

Linkage Analysis in the Families

Pairwise estimates of the recombination fraction (θ) and the corresponding LOD score (*Z*) between PDDR and each of the markers were calculated by using the MLINK computer program from the LINKAGE package (Lathrop et al. 1985). The COL2A1 and 3MH haplotypes consisting of *Hin*fI and *Pvu*II polymorphisms (Labuda et al. 1992) and tightly linked anonymous markers D12S14, D12S17, and D12S6 (Labuda et al. 1990) were run each as a single locus for linkage against PDDR.

Physical Mapping of the PDDR Locus

DNA from YAC clones (Research Genetics) were typed with microsatellites as described above, by using total yeast and human genomic DNA as negative and positive controls, respectively.

Demographic Model

Colonization of Ch-SLSJ (Jetté et al. 1991) was divided into seven migration waves ($j = 1,2, \ldots ,7$, starting at 1675, 1700, . . . , 1825), each within a time interval of 25 years corresponding to a single generation. Under the assumption of a uniform growth rate e^d per generation, the contribution of each immigration wave j to the total population N_t at time t is described by $N_{j_t} = N_{j_0} e^{d(t-t_{j_0})/25}$ (thus $N_t = \Sigma N_{j_t}$), where t_{j_0} describes the starting year of immigration of the founders within wave j and N_{j_0} is their number; $(t - t_{j_0})/25$ describes the number of generations g between the immigration date and time t. On the basis of knowing N_{j_0}'s between 1675 and 1850 and N_t at $t = 1850$ (Jetté et al. 1991), $d = .8$ was calculated from the equation above, by iteration (see fig. 1). If the PDDR carriers followed the same growth, their number in 1975, calculated from the same equation by assuming a single founder, was $N_g = e^{dg} \sim 15{,}000$ ($g = 12$ since 1675). This is consistent with the number of carriers, 12,000, in the present population of Ch-SLSJ (on the basis of the $1/26$ carrier rate). Historical records are less detailed for Acadians whose first settlement started at the beginning of 17th century; we used $g = 15$ as the putative historical age of the founder effect and $d = .6$, on the basis of the population growth between 1800 and 1951 (see the Introduction). Here again, $N_g = e^{dg} \sim 8{,}000$ is consistent with the present-day carrier population of 6,000, based on the $1/60$ carrier frequency.

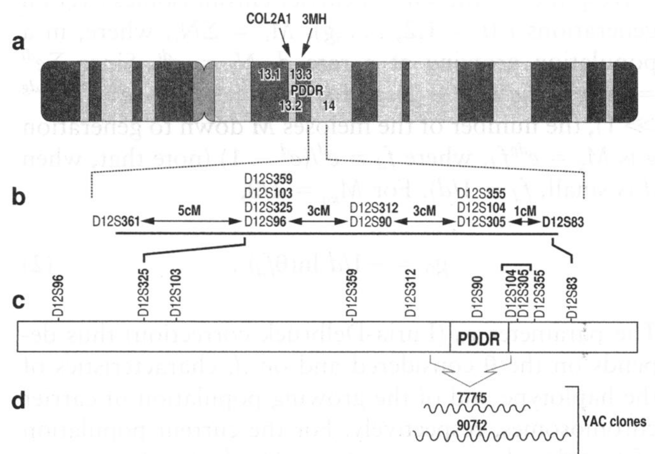

Figure 2 Steps in PDDR mapping. *a*, Localization of PDDR on human chromosome 12q13-14 (Labuda et al. 1990). *b*, Regional microsatellites map (Gaypay et al. 1994), with markers of unresolved order shown in columns. *c*, Refined map of PDDR and its flanking markers, based on LD and family studies. *d*, YAC clones positive for PDDR-flanking markers.

LD

LD was easily revealed by simple inspection of the data, because of the sufficient complexity of the tested haplotype. All noncarrier chromosomes were different from those carrying the disease, and the disequilibrium coefficient, D', was 1. $D' = (p_{cf} - p_c p_f)(p_c p_f p_{-c} p_{-f})^{-1/2}$, where p_{cf} is the frequency of the founder haplotype on carrier chromosomes, p_f is that of the founder haplotype, p_c is that of the carrier chromosomes, $p_{-c} = 1 - p_c$ is that of all the noncarrier chromosomes, and $p_{-f} = 1 - p_f$ is that of all the haplotypes other than the founder haplotype. Thus, when $p_{cf} = p_f = p_c$, $D' = 1$.

Quantitative analysis of LD data is based on the earlier works of Jennings (1917) and Robbins (1917). If loci recombine at a rate of θ per generation (i.e., genetic distance is θ), then a proportion $(1 - \theta)$ cosegregates, and, after g generations, a proportion $P = (1 - \theta)^g$ is expected to have cosegregated. At a small θ, $P \sim e^{-\theta g}$. Thus, knowing P from typing experiments, we can estimate either g or θ from the relation

$$\ln P = -\theta g \,. \quad (1)$$

However, in rapidly growing populations, according to the model of Luria and Delbrück (1943; also see Hästbacka et al. 1992) the recombinations immediately following the founder event are less likely than those during later generations. To estimate the likely number of recombinants (see eqq. 6–8 in Luria and Delbrück 1943), one omits the contribution of a number g_0 of generations immediately following the founder event, during which the number of meioses M is such that only one recombination is expected; that is, at the genetic distance θ, $M_{g_0} = 1/\theta$.

The number of meioses M since the founder event corresponds to the sum of carrier chromosomes over all generations i ($i = 1,2, \ldots ,g$), $M_g = \Sigma N_i$, where, in a population growing at a rate d, $N_i = e^{di}$. Since $\Sigma e^{di} = (e^{dg} - 1)e^d/(e^d - 1)$ and since $e^{dg} - 1 \sim e^{dg}$ (at $e^{dg} \gg 1$), the number of the meioses M down to generation g is $M_g = e^{dg} f_d$, where $f_d = e^d/(e^d - 1)$ (note that, when d is small, $f_d = 1/d$). For $M_{g_0} = 1/\theta$,

$$g_0 = -1/d \, \ln(\theta f_d) \,. \quad (2)$$

The parameter g_0 (Luria-Delbrück correction) thus depends on the θ considered and on d, characteristics of the haplotype and of the growing population of carrier chromosomes, respectively. For the current population of founder chromosomes $N_g = N_{g_0} e^{d(g-g_0)}$, thus $g - g_0 = (1/d) \ln(\theta N_g f_d)$. We will refer to the difference $g - g_0$ as g apparent, g_{app}, since this, rather than the historical g, is the likely value that is obtained from the experimental data. To calculate genetic distances θ from the observed P, g in equation (1) should be replaced by g_{app}:

$$\ln P = -\theta d[\ln(\theta N_g f_d)] \,, \quad (3)$$

and the range θ_- and θ_+ can be obtained from the equations

$$\ln P = -\theta_-/d[\ln(\theta_- N_g f_d) + 2] \,, \quad (3a)$$

$$\ln P = -\theta_+/d[\ln(\theta_+ N_g f_d) - 2] \,, \quad (3b)$$

respectively (Hästbacka et al. 1992; Kaplan et al. 1995).

Results

Linkage Studies in Families

PDDR families were typed with 12 microsatellite markers from the chromosome 12q13-14 region (fig. 2*b*). Significant Z values in two-point linkage analysis with the disease (table 1) were obtained for 10 of the microsatellites and for markers studied previously (Labuda et al. 1990). Recombinations critical for establishing the closest flanking markers were observed in the families shown in figure 3. In family 19, two affected sibs differed at the marker loci D12S305, D12S104, D12S355, and D12S83, revealing a maternal (I-1) recombination that placed these markers distal to the disease locus. In family 9, a crossover in individual I-2 (and detected in II-2) localized markers D12S355 and D12S83 distal to the PDDR locus as well as to both D12S305 and D12S104, separating the previously unordered loci (D12S305,D12S104) and D12S355. On the proximal side of PDDR, several recombinations with the cluster (D12S96,D12S325,D12S103,D12S359) were observed. Proximal location of D12S312 to the disease was established from a paternal (I-2) recombination in family 11, detected in affected individual II-4. Moreover, in family 9a, a maternal (II-1) recombination observed in unaffected individual III-4 separated proximal markers (D12S96, D12S325, D12S103, and D12S359) from the D12S312 locus, placing the latter closest to the disease. In summary, our family studies are consistent with the following locus order: cen-D12S361–(D12S96, D12S325,D12S103,D12S359)–D12S312–(D12S90, PDDR)–(D12S305,D12S104)–D12S355–D12S83–tel.

Haplotypes in Populations

We reconstructed all haplotypes consisting of five markers—D12S90,D12S305,D12S104,D12S355, D12S83—and established the phase of all but eight carrier haplotypes consisting of 10 markers (fig. 4). In the Ch-SLSJ population, the 11-5-4-5-2-4-7-1-1-9 consensus (haplotype 1) was unambiguously identified on 39 carrier chromosomes and was inferred on 2 others (as indicated by square brackets in fig. 4, haplotypes 4 and 10); 23 of these were independently sampled. All but one (haplotype 14) of the remaining Ch-SLSJ variants were identical with the consensus, at five or more loci

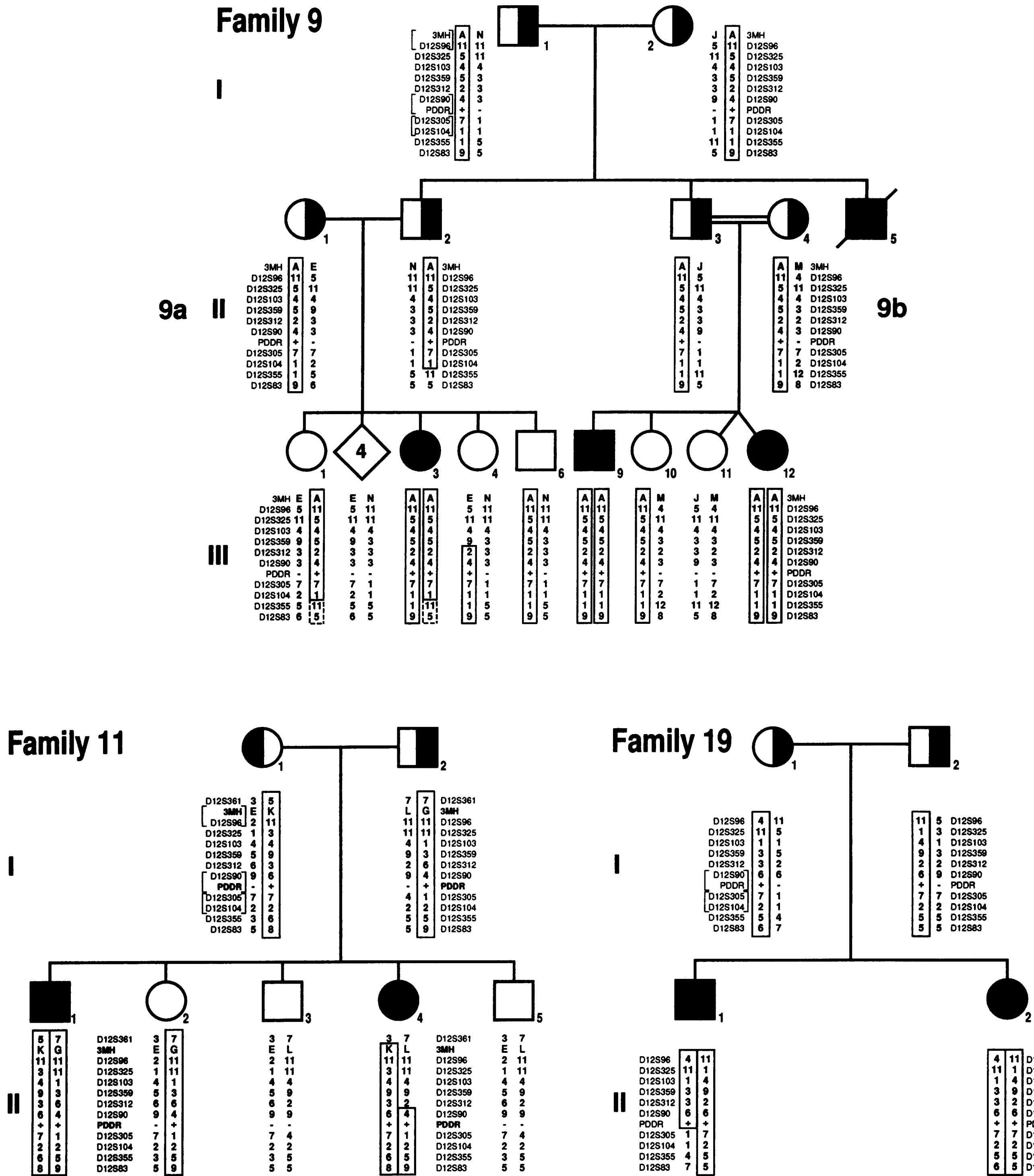

Figure 3 Crossovers near the PDDR locus. Disease and normal alleles are designated by plus signs (+) and minus signs (−), respectively. The numbers denote marker alleles, and brackets indicate ambiguity in the order of markers; and carrier haplotypes are boxed.

(indicated by black bars in fig. 4), indicating a single founder effect. In a much smaller Acadian group, the consensus haplotype (number 19) was 3-1-4-9-2-6-7-2-5-6, pointing to a single albeit distinct founder effect. In contrast, of the 35 haplotype variants on normal Ch-SLSJ and Acadian chromosomes (not shown), all were different. In two of three Polish families, PDDR patients were homozygous at either 10 or 5 loci of the studied haplotype. Similarly, a non-Acadian patient from maritime Canada was homozygous at nine loci. As expected,

Am. J. Hum. Genet. 59:633–643, 1996

PDDR (between D12S312 and D12S305)

		D12S96	D12S325	D12S103	D12S359	D12S312	D12S90	D12S305	D12S104	D12S355	D12S83	# affected chromosomes	# affected independent chromosomes
Founding chromosome	1	11	5	4	5	2	4	(7	1)	1	9	39	21
SLSJ Charlevoix	2	11	5	4	5	2	4	7	1	1	6	2	1
	3	11	5	4	5	2	4	7	1	5	5	2	1
	4	[11]	[5]	[4]	5	[2]	4	7	1	1	9	1	1
	5	[11]	[11]	[1]	5	[3]	4	7	1	1	9	3	2
	6	11	3	1	5	6	4	7	1	1	9	1	1
	7	2	5	4	5	2	4	7	1	1	9	1	1
	8	4	5	4	5	4	4	7	1	1	9	1	1
	9	5	11	1	5	3	4	7	1	1	9	1	1
	10	[11]	[5]	4	[5]	2	4	7	1	1	9	1	1
	11	[1]	[1]	1	[6]	4	4	7	1	1	9	2	2
	12	[4]	[2]	4	9	2	4	7	1	1	9	2	1
	13	[11]	[11]	4	5	3	4	7	1	1	9	4	2
	14	1	5	4	[3]	3	4	7	1	5	9	2	1
	15	11	11	4	4	2	4	7	1	1	9	1	
	16	11	3	1	9	2	4	7	1	1	9	1	
	17	11	5	4	5	2	4	7	1	1	5	1	
	18	11	5	4	5	2	4	7	1	11	5	1	
												total 66	37
Founding chromosome	19	3	1	4	9	2	6	7	2	5	6	3	3
Acadia	20	11	1	4	9	2	6	7	2	5	6	1	1
	21	11	1	4	9	2	6	7	2	5	5	2	1
	22	3	3	4	3	3	6	7	2	5	6	1	1
	23	4	1	4	9	2	6	7	2	5	6	2	1
	24	2	3	4	3	3	6	7	2	5	6	1	1
	25	11	11	4	[9]	2	6	7	2	5	6	2	1
	26	4	11	1	3	3	6	7	2	5	6	1	1
	27	4	11	1	3	3	6	1	1	4	7	1	
	28	11	11	1	9	6	6	7	2	5	6	1	
	29	11	11	4	9	2	6	7	2	5	5	1	
												total 16	10

Figure 4 PDDR-carrier haplotypes from Ch-SLSJ and Acadia. Haplotypes composed of 10 loci indicated at the top are shown in rows, with numbers denoting alleles. Square brackets indicate phase uncertainty. The phase of haplotypes 4 and 5 inferred in one family was unambiguously assigned in others, as was similarly done for haplotypes 10 and 11. Square brackets on haplotype 14 and 25 are single because of the separation of chromosomes from a patient with inferred phase into two groups of founders. The black bars denote segments identical to the ancestral haplotype; the gray bars denote segments recombined in previous generations; and the white bars denote segments recombined in the families studied. For a given haplotype variant, the total number of affected chromosomes and that of the independent chromosomes are shown to the right (we assume that chromosomes were independent if they went through six independent meioses; those related, if different, are also included in this group).

consanguinity still plays a role in the appearance of homozygous affected individuals in populations where overall incidence is very low.

As shown in figure 4, the haplotype D12S90-D12S305-D12S104 is the most conserved (4-7-1 in Ch-SLSJ and 6-7-2 in Acadia). The Acadian variant 6-7-2 was also identical to that in a Polish family (family 11; fig. 3). More interesting, the 9-5-2 variant found in another Polish patient was the same as both the non-Acadian PDDR haplotype from maritime Canada and yet another one from a family of European descent in Florida. Since we expect variants 6-7-2 and 9-5-2 to occur in the general population at frequencies of 7.3% and 1.6%, respectively, their presence on carrier chromosomes from different ethnic backgrounds is unlikely to be fortuitous and suggests identity by descent due to earlier common founder event(s).

Founder Effects and Demographic Model

The occurrence of founder effects for PDDR in French Canadians (in Acadia and Ch-SLSJ) should be considered in the context of this population's demographic history. Assuming a uniform growth rate, we estimated the contribution of each of the seven immigration waves between 1675 and 1850 to the total population of Ch-SLSJ (fig. 1). On the basis of this simple model (see Subjects and Methods) it appears that in the middle of 19th century the genetic contribution of the descendants of the first two immigration waves was >50%, a contribution that rises to >75% if the third wave is included. Because of the high growth rate (d = .8), after two to three generations the Charlevoix population had become sufficiently large ($N_{j_t} \gg N_{j_0}$) that the newcomers could not significantly alter its genetic constitution. The high incidence of a number of genetic disorders (including PDDR) with a carrier frequency of 1.5%–5% (see the Introduction) in the Ch-SLSJ region can thus be explained by rapid population growth after random sampling of chromosomes bearing rare mutations during the initial bottleneck (the term "bottleneck" applies to rare variants such as disease mutations but not necessarily to common polymorphisms of frequency $\geqslant$.05). This leads to the most likely figure of 12 generations (between 1675 and 1975) for the historical age of the founder effect in Ch-SLSJ and is consistent with genealogical studies (Bouchard et al. 1990; Heyer and Tremblay 1995), in the sense that genealogies of carriers can be traced back to the first settlers, whereas the same seems impossible with the founders from later generations. The same model is likely to apply for Acadia (see Subjects and Methods), for which we used 15 generations and a d value of .6. The existence of two different founder effects in Ch-SLSJ and Acadian populations is considered fortuitous but not surprising, since they derive from different groups of settlers.

Quantitative Consideration of the Haplotype Data

The probability of a recurrent recombination, recreating an ancestral carrier haplotype from an already recombined one, is negligible except for the events involving terminal loci of the haplotype when it is occupied by a frequent allele. In the latter case the terminal locus typed can be removed from the quantitative analysis, as was done with the leftmost D12S96 marker (allele 11, present on 65% of normal chromosomes).

The proportion P of the nonrecombined segments of the nine-marker haplotype, determined separately for

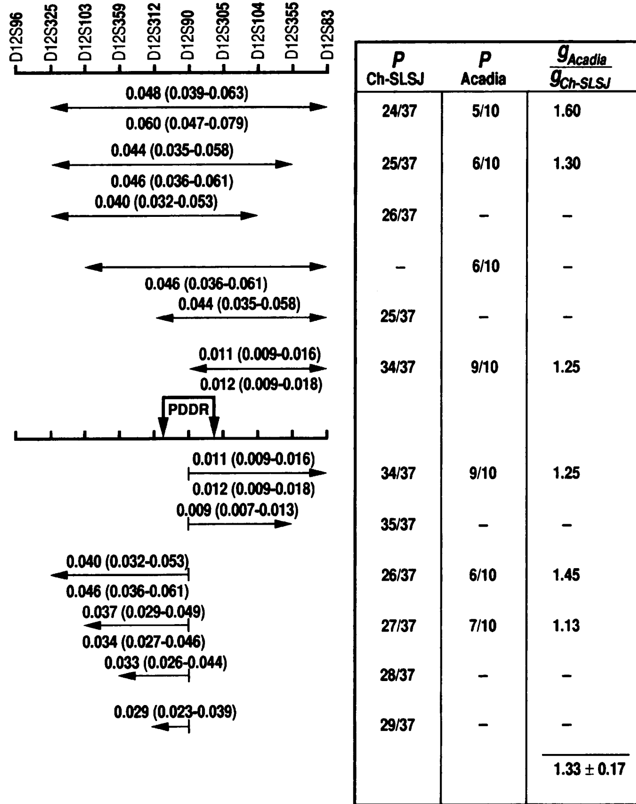

Figure 5 "Absolute" genetic distances between markers and between markers and the disease. The distances, in $\theta(\theta_-; \theta_+)$, above and below the arrows are for Ch-SLSJ and Acadia, respectively, and were calculated by using equation (3) (N_g values used were 15,000 for Ch-SLSJ and 8,000 for Acadia). P values denote the fraction of a nonrecombined haplotype segment (as indicated by arrows) calculated for independent chromosomes from figure 4. The rightmost column represents the age ratio of founder effects (g_{Acadia}:$g_{\mathrm{Ch\text{-}SLSJ}}$ = $\ln P_{\mathrm{Acadia}}$:$\ln P_{\mathrm{Ch\text{-}SLSJ}}$).

Ch-SLSJ and Acadian chromosomes, is given in figure 5. Using equation (3), which implements the Luria-Delbrück model in LD mapping (Luria and Delbrück 1943; Hästbacka et al. 1992), we calculated the "absolute" genetic distances between markers or between markers and the disease, within the studied haplotype. The θ values obtained for Ch-SLSJ and Acadia are in very good agreement (in spite of the small size of the latter data set), which supports the genetic model as well as the demographic and historical parameters used in the calculations.

Although the relative distances within the haplotype can be determined by comparing the corresponding ln$P$ values from chromosomes of the same founder group, the comparison of ln$P$ values for the same haplotype segment but from different founder effects provides information on their relative age. For example, when such pairwise combinations of P_{Acadia} and $P_{\mathrm{Ch\text{-}SLSJ}}$ for six haplotype segments are used, the mean ratio of g_{Acadia}:$g_{\mathrm{Ch\text{-}SLSJ}}$, 1.33, was obtained, consistent with the likely 15:12, or 1.25, historical ratio (fig. 5).

Finally, under the assumption of a single recombination between the disease and markers D12S305 and D12S104 among 37 Ch-SLSJ chromosomes, the distance estimated from equation (3) thus could be .005 (range .004–.007). The corresponding figure of .004 (range .003–.005) is obtained if the data from both populations are combined (47 chromosomes and Ch-SLSJ parameters). The haplotype segment for loci D12S90, D12S305, and D12S104 would be <0.4 cM. The fact that the variants 4-7-1 and 6-7-2 of this haplotype are shared by PDDR chromosomes of different ethnic origin is consistent with the old precolonial founder effect(s), since a segment of 0.4 cM is expected to remain unrecombined on half of the chromosomes after >200 generations. The genetic distance between the disease and the distal flanking markers estimated from the LD analysis being ~10 times shorter than that obtained by the family studies (table 1) encouraged us to undertake the experiments correlating the genetic and physical assignment.

Typing YAC Clones

Three human YAC clones from the Cohen library (Cohen et al. 1993) were tested for the presence of markers linked to the PDDR locus (table 2). Two of them, 777-f-5 and 907-f-2 (molecular weight 580 and 1,580 kb, respectively), were found to carry the proximal (D12S312) and distal flanking markers (D12S305,D12S104) as well as the marker D12S90 and thus, most likely, also the PDDR gene.

Discussion

Quantitative LD has been used in the genetic mapping of founder human populations (Hästbacka et al. 1992; Lehesjoki et al. 1993; Ellis et al. 1994; Sulisalo et al. 1994; Mitchison et al. 1995; Levy et al. 1996; Varilo et al. 1996). The Finnish founder population was considered ideal for such an approach, because of its size, age, and numerous hereditary disorders due to a founder effect (de la Chapelle 1993). Here we show that popula-

Table 2

PCR Amplification of PDDR-Linked Markers on YAC Clones

	777-f-5	790-f-7	907-f-2
D12S103	−	+	−
D12S359	−	−	−
D12S312	+	+	+
D12S90	+	−	+
D12S305	+	−	+
D12S104	+	−	+
D12S355	+	−	+

NOTE.—Plus signs (+) and minus signs (−) denote presence and absence of marker-locus amplification, respectively.

tions as young as those of Acadia and Ch-SLSJ in Quebec are also suitable for LD mapping to narrow down genetic distances of tightly linked markers. Although the information on the markers flanking the disease came from the family studies, only the insight into historical meioses allowed us to refine distances between the disease and the distal markers D12S305,D12S104. Marker D12S90, which did not show any recombination with the disease locus, either in the historical meioses or in the families analyzed, is likely to be even closer. Finding the same variants (4-7-1 and 6-7-2) of these three-marker haplotypes in patients from different European groups not only indicates an ancient precolonial origin of the founding mutation(s) but also supports the tight linkage of the distal markers D12S305, D12S104, and D12S90 with the disease. A much more remote proximal flanking marker (D12S312) is also found on the same YAC clone, but, in light of our LD data, the disease gene is likely to reside on the distal marker's side.

Using a simple demographic model (fig. 1), we have shown that the founder effect of a number of genetic disorders in Ch-SLSJ can be explained by (*a*) sampling rare mutations due to a limited number of the first settlers at the turn of the 17th century and (*b*) the important genetic contribution of the first settlers to the following generations. Given the size of the population in the middle of the 19th century (fig. 1), we do not expect any significant changes in the genetic pool of Ch-SLSJ during its subsequent expansion by the end of the 19th century and in the 20th century. However, stochastic fluctuations, such as those that have been documented by genealogical studies (Heyer 1995), could have been of great importance at the beginning of the population. Among the first few settlers, some could have contributed more than others to the genetic pool, thus explaining the elevated carrier rate of some hereditary disorders (see the Introduction) and the absence of others that are frequent outside the region (e.g., Friedreich ataxia and Tay-Sachs disease, which do not exist in SLSJ, are prevalent in southeastern Quebec [Harding 1984; Gravel et al. 1995]).

The knowledge of the age of the founder effect is required for calculation of "absolute" θ's from equation (1). Use of the historical $g = 12$ in Ch-SLSJ to calculate the length of the total haplotype from $P = 24/37$ in Ch-SLSJ yields $\theta = .036$ (the corresponding value for Acadia is $\theta = .046$, for $g = 15$). In contrast, if $\theta = .07$, known independently from Généthon linkage studies on CEPH families (Gyapay et al. 1994), is used to calculate the age of the founder effect from the same data, 6.2 generations are obtained for Ch-SLSJ (9.9 for Acadia). Accepting this number would place the origin of the founder effect at the first half of the 18th century, when the population of Charlevoix and, subsequently, that of Ch-SLSJ, already numbered 4,000–18,000 individuals. This is inconceivable with the present-day carrier rate of 1/26, as discussed above. This discrepancy is explained by the fact that equation (1) does not apply to the growing population (Luria and Delbrück 1943) and hence leads to the underestimation of the θ value (tantamount to the underestimation of the age if θ is independently known). The problem is resolved by replacing g in equation (1) by g_{app} in equation (3). The correction of the age of the founder effect can be obtained from equation (2): $g_0 = 2.6$ ($\theta = .07$, $d = .8$) for CH-SLSJ, and $g_0 = 3.1$ ($\theta = .07$, $d = .6$) for Acadia. These g_0 values, when added to the experimental g_{app} above, lead to 8.8 and 13 generations, respectively, approaching the historical estimates. The Luria-Delbrück model thus provides a better description of experimental data than does equation (1). However, it requires the knowledge of additional parameters, the growth rate d, and the number N_g of carrier chromosomes in the present-day population. Usually there will be enough historical knowledge to estimate a realistic range of d values to be considered.

In the likelihood estimation of Kaplan et al. (1995), the parameters d (therein denoted "λ") and g are treated as a nuisance confounded in the size of the disease population (as our N_g above); this is especially convenient when these parameters are impossible to estimate. However, the use of demographic parameter $\lambda \ll 1$ in the modeling of the Poisson branching process by Kaplan et al. (1995) excludes our data set where large d values (.8 and .6) are observed. Application of this approach to the data on Finnish diastrophic dysplasia leads to θ values larger than those obtained within the framework of Luria-Delbrück formalism. This would indicate that our corrected age $g_{app} + g_0$ is an underestimate—that is, that the g_0 correction (eq. 2) is conservative. On the other hand, Levy et al. (1996) used both the Luria-Delbrück and Kaplan et al. (1995) models to analyze the data on familial Mediterranean fever in Moroccan Jews and found good agreement between the corresponding estimations of θ. Larger bounds in the approach of Kaplan et al. (1995) could be explained by the fact that both confounded parameters g and λ are adjustable, whereas in the Luria-Delbrück model d is fixed. However, by selecting a range of d values, one would automatically obtain the range of evaluated distances. Alternatively, one may extend the lower and upper estimates of θ by using a range of g at a given d, as did Varilo et al. (1996), who used the likelihood approach of Terwilliger (1995) to analyze linkage disequilibrium in the CLN5 locus.

Given the uncertainty of the genetic distance:physical distance ratio, the use of equation (1) rather than equation (3) can be, in some cases, of little consequence. An average correspondence of 1 cM to 1 Mb underestimates physical distances close to the centromeres and overestimates them in telomeric regions (Watkins et al. 1994); for example, 1 cM over 17 Mb was reported on human

9p13-21.1 (Kwiatkowski et al. 1993). Such deviations from a 1 cM:1 Mb genetic distance:physical distance ratio also could be responsible for our finding of the marker D12S312, separated from the distal markers by ~3 cM, on the same YAC clone.

Realizing the difference between historical g and g_{app} is essential if we want to evaluate the age of the founder effect and to understand the demographic and genetic history of a population. In spite of its limitations, the advantage of Luria-Delbrück formalism is that it provides a simple formula for estimating the genetic distances and the age of the founder effect. For assessing the magnitude of the age correction required in a growing population, the value of g_0 can be easily calculated for a range of θ and d. The application of the Luria-Delbrück correction to recently reported results on the founder effect for idiopathic torsion dystonia (ITD) in Ashkenazi Jews (Risch et al. 1995*b*) illustrates this point. ITD mutation has been estimated as having originated between 1400 and 1750—that is, when the Jewish population in eastern Europe already numbered between ⩾10,000 (in 1400) and ~450,000 (in 1750), respectively (Beinart 1992; Motulsky 1995). The above dating places the mutation frequency at ⩽1/10,000 at the start—and most likely at 1/100,000—contrasting with the present-day frequency of 1/2,000–1/6,000. Since, however, the dating corresponds to the apparent number of generations, the postulated age of the founder effect should be increased. On the basis of the length of the haplotype (θ = .018) linked to ITD used in the study by Risch et al. (1995*b*) and assuming that d = .3 (Risch et al. 1995*a*), we estimated the Luria-Delbrück correction, $g_0 \sim 9$. This correction, which, as we discussed above, is likely to be conservative, moves the origin of the founder effect back in time by ⩾2 centuries, to the Middle Ages. The ITD founder event would thus coincide with the founding of the Jewish population in eastern Europe, and the same could be the case for a number of genetic diseases known to be endemic to this population (Motulsky 1995). The origin of the founder effect in the early 12th century would be consistent with a demographic model similar to that of Ch-SLSJ, in which the first settlers contributed much more to the future generations than did the later immigrants. Historical data, such as either coins with Hebrew inscriptions minted under the reign of Mieszko III the Old at the end of 12th century or the privileges granted to Jews in the Statue of Kalisz in 1264 by Boleslaw V the Pious (Davies 1984; Beinart 1992), document early settlements of Ashkenazim in medieval Poland. Subsequent migration eastward followed by great demographic expansion was presumably promoted by Polish-Lithuanian union starting by the end of the 14th century and putting the territories of present-day Lithuania, Belarus, Ukraine, and Poland under a single Jagiellonian rule. The frequent occurrence of a number of otherwise rare genetic disorders thus could be often explained by migration and demographic expansion, such as that in Finns (de la Chapelle 1993) and French Canadians.

Acknowledgments

We are grateful to Dr. Chiara Prinster for participating in some experiments and to Dr. Wagner Yotow for discussions. We thank J. Wishart for illustrations and E. Wan for oligonucleotide synthesis. This work was supported by the Shriners of North America. D.L. acknowledges support (a research grant) from the Medical Research Council of Canada.

References

Arsenault B (1978) Histoire des Acadiens. Leméac, Ottawa

Beinart H (1992) Atlas of medieval Jewish history. Simon & Schuster, New York

Bodmer WF (1972) Population genetics of the HL-A system: retrospect and prospect. In: Dausset J, Colombani J (eds) Histocompatibility testing. Munksgaard, Copenhagen, pp 611–617

Bouchard G, Roy R, Casgrain B, Bourque M (1990) L'état civil saguenayen et la reconstitution automatique des familles à l'aide du système SOREP. In: Bouchard G, De Braekeleer M (eds) Histoire d'un génome. Presses de l'Université du Québec, Quebec, pp 21–46

Canadian Encyclopedia, The (1984) Hurting Publishers, Edmonton, vol 1, pp 5–6

Cohen D, Chumakov I, Weissenbach J (1993) A first-generation physical map of the human genome. Nature 366:698–701

Davies N (1984) God's playground: a history of Poland. Columbia University Press, New York

De Braekeleer M (1991) Hereditary disorders in Saguenay-Lac-St-Jean (Quebec, Canada). Hum Hered 41:141–146

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

Delvin EE, Glorieux FH, Marie PJ, Pettifor JM (1981) Vitamin D dependency: replacement therapy with calcitriol. J Pediatr 99:26–34

Ellis NA, Roe AM, Kozloski J, Proytcheva M, Falk C, German J (1994) Linkage disequilibrium between the *FES, D15S127,* and *BLM* loci in Ashkenazi Jews with Bloom syndrome. Am J Hum Genet 55:453–460

Fraser D, Kooh SW, Kind PH, Holock MF, Tanaka Y, DeLuca HF (1973) Pathogenesis of hereditary vitamin D-dependent rickets. N Engl J Med 289:717–782

Gahl WA, Schneider JA Aula PA (1995) Lysosomal transport disorders: cystinosis and sialic acid storage. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 3763–3797

Gauvreau D, Guérien M, Hamel M (1991) De Charlevoix au Saguenay: mesure et caracteristiques du mouvement migratoire avant 1911. In: Bouchard G, De Braekeleer M (eds) Histoire d'un génome. Presses de l'Université du Québec, Quebec, pp 145–162

Gravel RA, Clark JTR, Kaback MM, Mahuran D, Sandoff K, Suzuki K (1995) The G_{m2} gangliosidoses. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 2859–2860

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nature Genet 7:246–339

Harding AE (1984) The hereditary ataxias and related disorders. Churchill Livingstone, Edinburgh

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

Heyer E (1995) Mitochondrial and nuclear genetic contribution of female founders to a contemporary population in northeast Quebec. Am J Hum Genet 56:1450–1455

Heyer E, Tremblay M (1995) Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. Am J Hum Genet 56:970–978

Jennings HS (1917) The numerical results of diverse systems of breeding, with the respect to two pairs of characters, linked or independent, with special relation to the effect of linkage. Genetics 2:97–154

Jetté R, Gauvreau D, Guérien M (1991) Aux origins d'une région: le peuplement fondateur de Charlevoix avant 1850. In: Bouchard G, De Braekeleer M (eds) Histoire d'un Génome. Presses de l'Université du Québec, Quebec, pp 75–106

Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56:18–32

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

Kornfeld S, Sly WS (1995) I-Cell disease and pseudo-Hurler polydystrophy: disorders of lysosomal enzyme phosphorylation and localization In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 2495–2508

Kwiatkowski DJ, Dib C, Slaugenhaupt SA, Povey S, Gusella JF, Haines JL (1993) An index marker map of chromosome 9 provides strong evidence for positive interference. Am J Hum Genet 53:1279–1288

Labuda M, Fujiwara TM, Ross MV, Morgan K, Garcia-Heras J, Ledbetter DH, Hughes MR, et al (1992) Two hereditary defects related to vitamin D metabolism map to the same region of human chromosome 12q13-14. J Bone Miner Res 7:1447–1453

Labuda M, Morgan K, Glorieux FH (1990) Mapping autosomal recessive vitamin D dependency type I to chromosome 12q14 by linkage analysis. Am J Hum Genet 47:28–36

Labuda M, Lemieux N, Tihy F, Prinster C, Glorieux FH (1993) Human 25-hydroxyvitamin D 24-hydroxylase cytochrome P450 subunit maps to a different chromosomal location than that of pseudovitamin D-deficient rickets. J Bone Miner Res 8:1397–1406

Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Lehesjoki AE, Koskiniemi M, Noria 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

Levy HL, Taylor RG, McInnes RR (1995) Disorders of histidine metabolism. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 1107–1123

Levy EN, Shen Y, Kupelian A, Kruglyak L, Aksentijevich I, Pras E, Balow JE Jr, et al (1996) Linkage disequilibrium mapping places the gene causing familial Mediterranean fever close to D16S246. Am J Hum Genet 58:523–534

Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511

Mitchell GA, Lambert M, Tanguay RM (1995) Hypertyrosinemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 1077–1106

Mitchison HM, O'Rawe AM, Taschner PEM, Sandkuijl LA, Santavuori P, de Vos N, Breuning MH, et al (1995), Batten disease gene, *CLN3:* linkage disequilibrium mapping in the Finnish population, and analysis of European haplotypes. Am J Hum Genet 56:654–662

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

Prader A, Illig R, Heierli E (1961) Eine besondere Form der primären Vitamin D resistenten Rachitis mit Hypocalcemie und autosomal dominantem Erbgang: die hereditäre Pseudomangelrachitis. Helv Paediatr Acta 16:452–468

Risch N, de Leon D, Fahn S, Bressman S, Ozelius L, Breakefield X, Kramer P, et al (1995*a*) ITD in Ashkenazi Jews—genetic drift or selection? Nat Genet 11:14–15

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

Robbins RB (1917) Some applications of mathematics to breeding problems. Part III. Genetics 3:375–389

Scott, CR (1995) Sarcosinemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of inherited disease, 7th ed. McGraw-Hill, New York, pp 1329–1335

Serre JL, Simon-Bouy B, Mornet E, Jaume-Roig B, Balassopoulou A, Schwartz M, Taillandier A, et al (1990) Studies of RLFP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. Hum Genet 84:449–454

Sulisalo T, Klockars J, Mäkitie O, Francomano CA, de la Chapelle A, Kaitila I, Sistonen P (1994) High-resolution linkage-disequilibrium mapping of the cartilage-hair hypoplasia gene. Am J Hum Genet 55:937–945

Tanaka KR, Paglia DE (1995) Pyruvate kinase and other enzyme deficiency disorders of the erythrocyte. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Stanbury JB, Wyngaarden JB, Fredrickson DS (eds) The metabolic and molecular basis of

inherited disease, 7th ed. McGraw-Hill, New York, pp 3485–3511
Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56: 777–787
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
Watkins WS, Zenger R, O'Brien E, Nyman D, Eriksson AW, Renlund M, Jorde LB (1994) Linkage disequilibrium patterns vary with chromosomal location: a case study from the von Willebrand factor region. Am J Hum Genet 55: 348–355