

# HUMAN POPULATION GENETICS: Lessons from Finland

---

Juha Kere

*Finnish Genome Center, University of Helsinki, Helsinki 00014, Finland;  
e-mail: juha.kere@helsinki.fi*

**Key Words** disease gene, genetic mapping, genetic marker, recessive inheritance, common disease

■ **Abstract** A population of about 5 million at the northern corner of Europe is unlikely to arouse the attention of the human genetics community, unless it offers something useful for others to learn. A combination of coincidences has finally made this population one that, out of proportion for its size, has by example shaped research in human disease genetics. This chapter summarizes advances made in medical genetics that are based on research facilitated by Finland's population structure. The annotation of the human genome for its polymorphism and involvement in disease is not over; it is, therefore, of interest to assess whether genetic studies in populations such as the Finnish might help in the remaining tasks.

## INTRODUCTION

### Finland as a Model Population for Human Genetics

Finland is inhabited by a population that is readily distinguished from other European populations by its unusual language, which does not belong to the large family of Indoeuropean languages. Finnish and three other major languages, Hungarian, Estonian, and Saame, stem from the Uralic language family. This geographically marginal population belongs to the genetically tight cluster of European populations (5).

Although the relationship of Finns to the other European populations has been the subject of many population genetic studies, less emphasis is often put on the internal structure of the population, even though it has current practical importance. Isolated founder populations have become emphasized as fields for the study of common diseases and genetic susceptibility (17, 42, 92). I believe that it is important to regard such research from a broad population point of view that could help to avoid many pitfalls and, thus, arrive at more likely interpretations of the data. Therefore, in this review I concentrate primarily on medical genetics, its

lessons about recent genetic history, and the structure of the present population, with particular attention to genes in disease.

## Geography and Climate

Some background information is useful to understand basic ideas on population structure. Finland spans from 59° to 70° northern latitude. This location puts Finland's capital, Helsinki, at the same latitude as Anchorage, Alaska. The climate of Finland is, however, much milder. In southern Finland, the average temperature in July is about 17°C, whereas the average winter temperatures remain below 0°C from December to March. Modern agriculture yields two crops of wheat and one crop of barley, oats, and rye, but no corn or rice. During recorded history, poor years have caused famines several times, the last one in the 1860s.

Finland has land boundaries with Russia, Norway, and Sweden, and a seashore that spans about a third of its circumference. The area of Finland is 338,000 sq km (131,000 sq mi), ranking in size just after Alaska, Texas, California, and Montana, or between Italy or Poland on one hand and Germany on the other. Most of the area is covered by forest (68%) or wetland (11%), and a large fraction by numerous lakes (9.9%), leaving 11% for agriculture or built-up area. The average population density is presently about 17 inhabitants per sq km; however, this is not a relevant figure because the population is very unevenly spread. The northernmost half is inhabited by 13% of the population, whereas 27% of the population reside in the six largest cities (79).

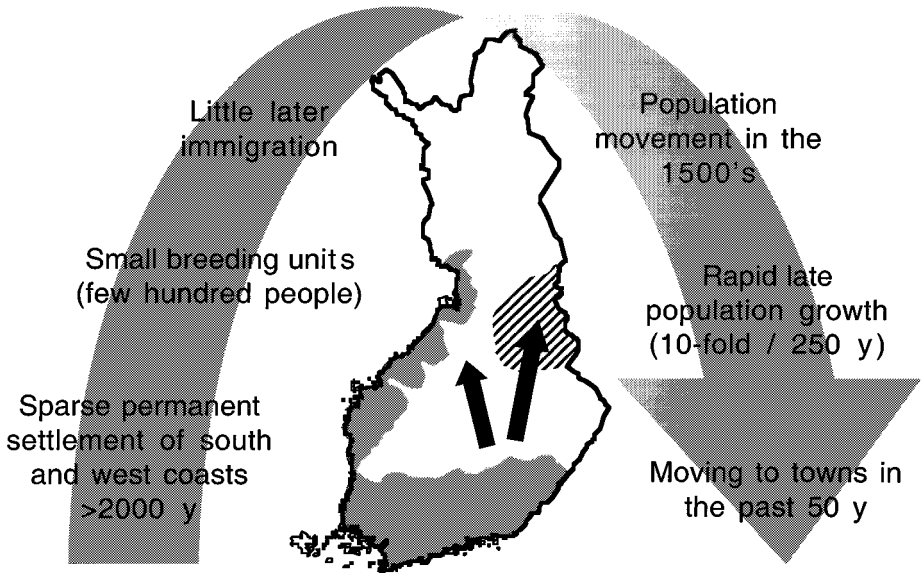
## Language as a Cultural Barrier

The Finnish language is completely unintelligible to all its neighbors except Estonians. Throughout historical times, the linguistic barrier has discouraged immigration, which has not exceeded the annual level of 0.2%, first reached in the 1960s. In addition to Finnish, native minority populations have Swedish (5.7%) or Saami (0.03%) as their mother tongues; other languages are spoken by 1.4% of the population (79).

Genetic admixture between the Swedish-speaking and Finnish-speaking populations has occurred, even though the linguistic cultures have remained distinct (88). The Saami population, residing presently in the northernmost part of Finland, represents a distinct population within Finland. It seems to have avoided admixture with the southern Finnish population in spite of the linguistic similarity and remains an outlier in the family tree of European populations (5, 35–37).

## Early History

Although archaeological signs of human activity date back more than 9000 years, the oldest archaeological signs of agriculture are from the end of the Stone Age, 3300 to 4000 years ago (59). The southern and western coastal regions have had permanent settlement for at least 2000 years, whereas vast parts of the country were inhabited permanently only after A.D. 1550 (Figure 1).



**Figure 1** A brief population history of Finland. Starting from lower left corner and moving clockwise, the main features of population history are listed. The map shows the coastal region settled permanently before A.D. 1550 (dotted area), and the arrows indicate population movement in the sixteenth and seventeenth centuries. The province of Kainuu is highlighted (striped area).

A question of much local interest concerns the origin of Finns, but an embraced understanding has not been achieved among archaeologists, linguists, and geneticists. The problems are obvious: Although archaeological findings provide indisputable evidence for settlement, it is often difficult to tell who the people were, i.e., to distinguish cultural transfer from population movement. A similar problem concerns linguistic interpretations. Genetic evidence pointing to a small relationship between the Finnish and Saami populations is not readily evident from their close linguistic relationship (74). Finally, the continuity of sparse populations in contrast to replacement by new populations is a difficult question to solve. Genetic evidence can seldom help with genes that were lost, thus biasing its view of history.

Summarizing much of the population genetic data from Europe, Cavalli-Sforza et al (5) place Finns closest to their geographical neighbors. Recent evidence from Y-chromosomal variation in Finland has indicated remarkably low diversity, suggesting a narrow bottleneck, whereas mitochondrial diversity is broader (29, 37, 75). A European founder effect may also have its origins in early Uralic populations. A single mutation has been found in the chemokine receptor 5 gene (CCR5) that confers protection against HIV-1 infections. In 18 European populations studied, the  $\delta$ -CCR5 variant is embedded in a haplotype that is otherwise rare.

The highest frequencies of the variant were found in the Finnish and Mordvinian populations (16%), and the lowest in Sardinia (4%), suggesting a northeastern origin (47).

## POPULATION STRUCTURE

### Isolation by Density

Before modern times, inhabitation concentrated around waterways—sea, lakes, and rivers—that allowed easier communication and trade. A patchwork of forests and wetland was rich in game but uninviting for traffic, even after large areas were cleared for agriculture. The population has been estimated at about 250,000 until the 1600s. Even with consideration for the large parts of the country that were uninhabited, the population density in those times was very low. Walking and paddling in summer and skiing in winter were the most common means of communication, which must have had strong social and, thus, genetic consequences.

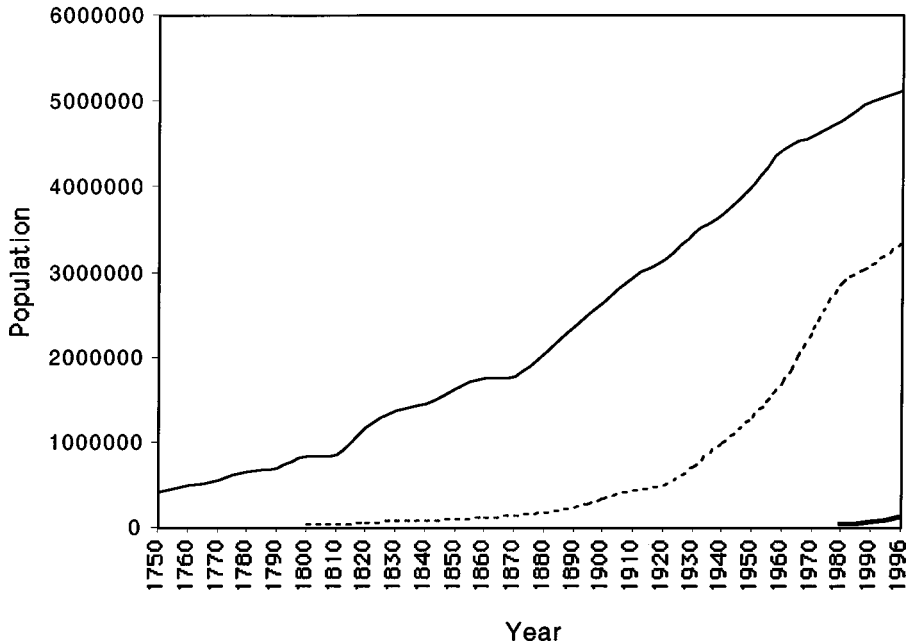
Indeed, mating probably occurred in units that consisted of hundreds, rather than thousands, of individuals; panmixis was not the word of the day. This concept has been called isolation by density, or dynamic isolation (56), and it refers to multiple small breeding units and little gene flow between the units. Small breeding units and little external gene flow allowed genetic drift to play with gene frequencies.

### Population Expansion

Much of the population spreading occurred in the seventeenth century, when the previous wildmarks were settled permanently (Figure 1). Finland's population has grown at an unusually high rate, approximately 10-fold, during the past 250 years (Figure 2). The growth rates in Finland through most of the 1800s are comparable to those of the fastest-growing populations of Asia today. The oldest census records available show that from 1751–1760, the population averaged 457,000, with 20,500 live births and 13,300 deaths annually, yielding a growth figure of 1.5% per year. By comparison, from 1991–1995, the same figures were 5,064,000; 65,100; 49,500; and 0.45%, respectively. The proportion of children <15 years was 37% in 1751 and 19% in 1995 (79). Until World War II, the population remained mostly rural, and there is direct evidence that the breeding units remained small, continuing the concept of dynamic isolation with the increased population density that was, however, still low in absolute terms.

Some direct evidence for breeding unit size comes from studies by Nevanlinna (56) who studied population registries to assess the geographical origins of marriages. His study of marital habits in Hirvensalmi (a community in central southern Finland) during two periods in the 1800s showed that half of the marriages were between individuals from the same village (each village had 100–500 inhabitants). Based on these and other observations, Nevanlinna estimated the breeding unit size

## Population of Finland 1750-1996



**Figure 2** Population growth in Finland. Reliable statistics are available from A.D. 1750 onward. The solid line shows total population, and the dotted line represents urban population. The bold line (bottom right corner) depicts the population born outside Finland.

at 200 to 600 individuals. Subsequent studies have been published from other communities with different geographical structures (73). Cousin marriages, however, have been remarkably rare, indicating active avoidance as compared to random mating (18). There is reason to believe that these patterns prevailed through much of the period of rapid population growth.

### Founder Effects and Drift

What should happen with genes in such circumstances? In a computer simulation using POPULUS software (V. Ollikainen & H. Mannila, unpublished), 50 founders were expanded to a population of 100,000 in 20 generations (corresponding to an annual growth rate of about 1.5%). The expanding population was allowed to mate randomly. The fate of each of the founding chromosomes was followed for one locus, and the effects of random drift (loss and enrichment of alleles) were recorded. The simulation was repeated 1000 times; the results are shown in Figure 3. Starting from a frequency of 1% for each allele (100 founding chromosomes), a rapid enrichment occurred for some alleles, reaching 6% for the most common allele

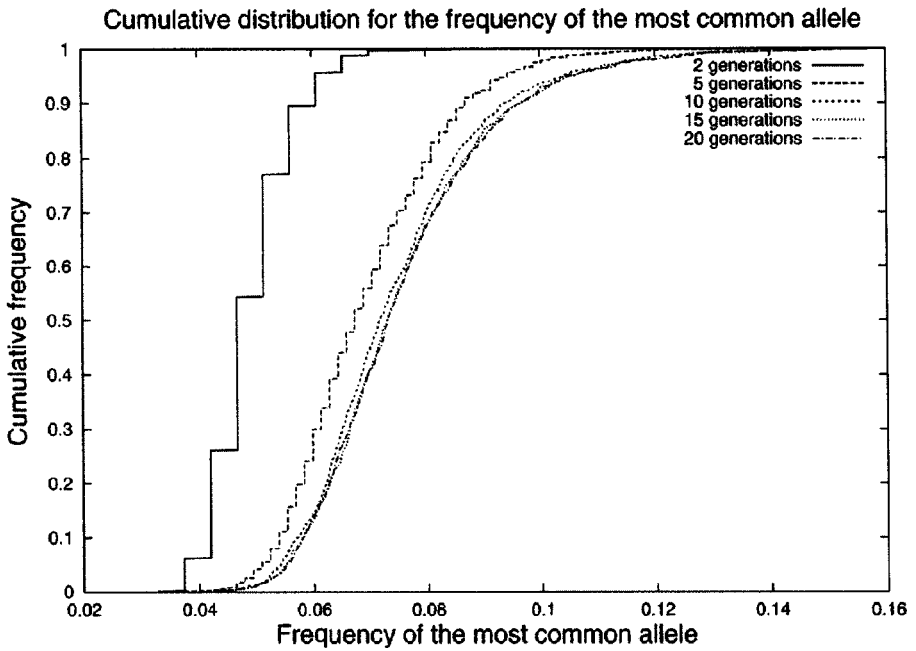
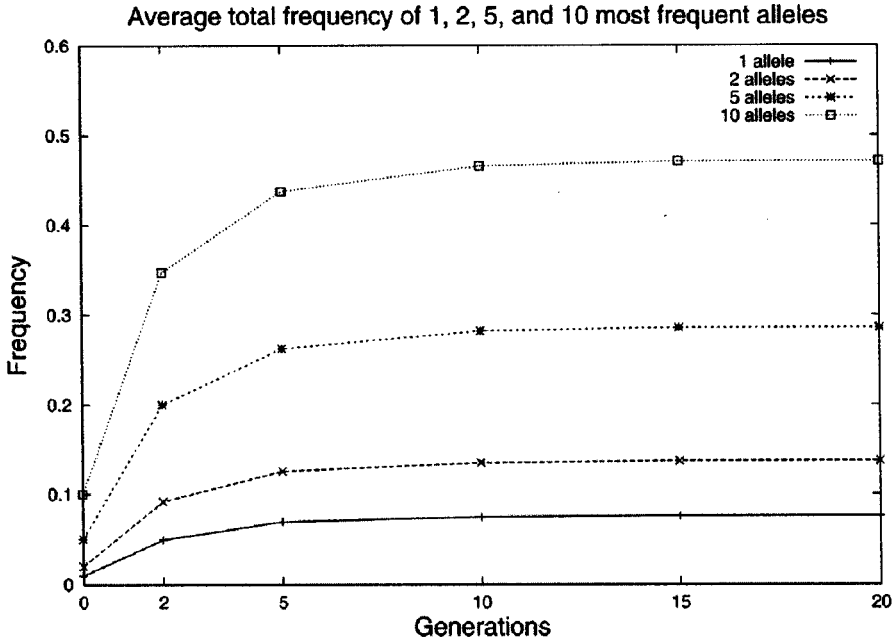
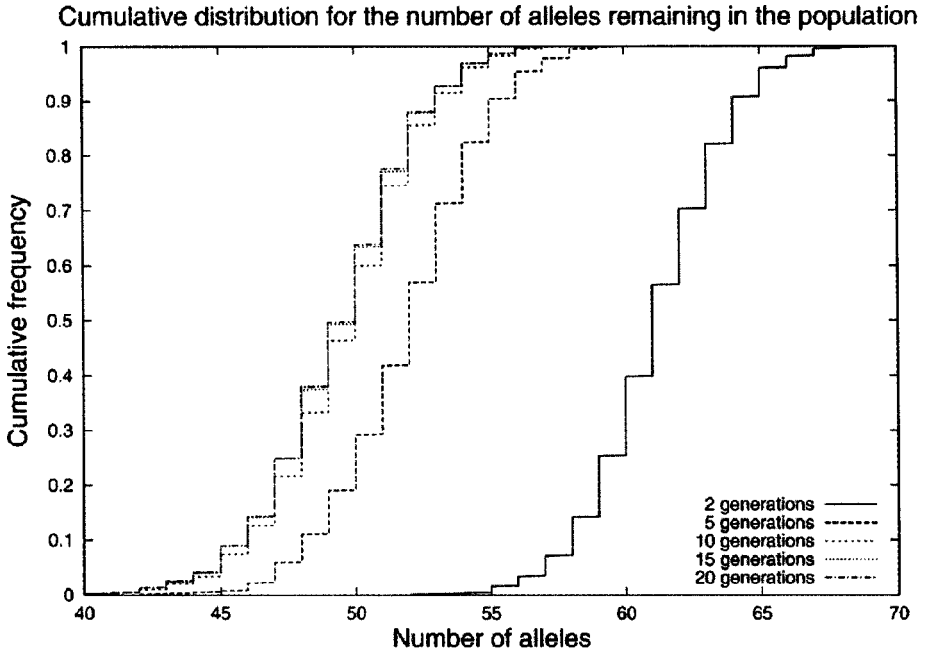


Figure 3 (Continued)



**Figure 3** Effect of drift in a small expanding population. A simulation was performed using the POPULUS software (V. Ollikainen & H. Mannila, unpublished). 50 founders (100 founding alleles) expanded to 100,000 over 20 generations. Fate of alleles was recorded at 2, 5, 10, 15, and 20 generations (population size 107, 334, 2236, 14,954, and 100,000, respectively). The top panel shows increase in frequency for the most common alleles (average from 1000 simulations). The middle panel shows the frequency distribution of the most common allele in 1000 simulations. In all simulations, the most common allele was enriched almost 5-fold and, though rarely, up to 15-fold. The bottom panel shows the number of alleles remaining in the population in 1000 simulations. Only 40%–57% of the founder alleles survived in the population.

and levels of over 40% for the 10 most common alleles. The enrichment was accompanied by the rapid loss of most alleles, so that 40% to 57% of alleles survived after 10 generations in different simulations (Figure 3). This simulation illustrates what strong effects random drift may have in small expanding populations.

### The Example of Kainuu

The province of Kainuu has become one of the model populations for genetic studies. It is located at the narrowest middle part of Finland, extending roughly from the midline to the eastern border and spanning about 200 km vertically (Figure 1). Approximately 95% of the total area is used for forestry. Kainuu is sparsely inhabited, with only 4.1 inhabitants/sq km.

The population structure and regional history of Kainuu have been extensively studied (24, 27). First signs of permanent settlement, attained by pollen analysis, date back to the eighth century, but the region remained largely uninhabited until the sixteenth century. Settlements were totally destroyed during the Russian war, 1570–1595, and resettlement continued until the mid-seventeenth century, mainly from the Savo region (75%) but also from Ostrobothnia (25%). Population data from those dates come from account books, and population has been estimated at 1400 to 2700 persons, in about 240 households, before the 1650s. After that period, church records are also available for accurate figures. Over the period 1880–1963, the population grew faster than in the rest of Finland, but since that time, emigration has led to a decrease in the population.

Immigration was low during the entire period of rapid growth, consisting in large part of return migration by people born in the region. The region is further characterized by extremely stable ecclesiastical and administrative boundaries that date back to 1599 (Paltamo parish). Throughout its history, the region has belonged to the same bishopric, province, court of appeal and of lower appeal, and inferior court (21). The eastern boundary still follows the line drawn after the Russian War in 1595. In 1996, the population in the region was 94,000, including the town of Kajaani (population 37,000).

## MOLECULAR GENETIC MAKEUP OF THE FINNISH POPULATION

### Common and Rare Alleles

Nevanlinna (56) used common and rare blood group antigens to study genetic drift and assess the genetic makeup of different communities within Finland. He chose seven communities from four non-neighboring locations and measured allele frequencies for 10 loci in carefully ascertained samples. The results show unmistakably the effect of genetic drift: Gene frequencies in each of the seven communities varied up to over six standard deviations around the mean. At the same time, the gene frequencies in individual communities distributed symmetrically around the national frequency estimates. These data were interpreted to indicate that individual communities had a common seed population from which the subisolates were formed by drift. Nevanlinna estimated further that about 5% of subisolates had undergone gene enrichments greater than six standard deviations (57). Similar wide geographical variations for allele frequencies have been observed for HLA antigens in a large set of 10,000 subjects (78).

For common alleles, the overall frequencies in Finland are similar to other European countries. This is most easily seen now for the numerous microsatellite marker alleles that have been used in various gene mapping studies elsewhere and in Finland. However, a closer look at small subisolates reveals variations that may be significant even though they are not as numerically dramatic as those for

rare genes. For example, the frequency of the common blood group allele A1, with an overall frequency of 0.213 in Finland, varied between 0.136 ( $-3$  SD) and 0.253 ( $+2.5$  SD) in the seven communities studied by Nevanlinna (56). At the province level, the gene frequencies for common alleles are already closer to the mean.

The small number of settlers and the small size of the breeding units caused a total loss of very rare alleles in most subisolates. Nevanlinna observed that the frequency of genes found in southern Finland at frequencies of 1/500 or less were completely absent from the parts of Finland that were settled later. On the other hand, genes with frequencies in the few percent range often showed higher than 10-fold differences in frequency among the seven different communities that he studied (56). The genes causing rare recessive Finnish diseases belong to this category and consequently show the typical uneven distributions for parental or grandparental birthplaces. In an ongoing study, we are using microsatellite markers and SNPs to assess the distribution of various rare alleles further (M.L. Savontaus, P. Lahermo, P. Sistonen, E. Salmela, & J. Kere, unpublished).

## Finnish Disease Heritage

Finland was called the “promised land of rare hereditary traits” by Norio, Nevanlinna, and Perheentupa (58), but the idea was better expressed in the term “Finnish disease heritage,” which is used to refer to the recessive diseases occurring in Finland at higher than usual frequencies (as compared to other European countries). The list of diseases now includes 35 entities (Table 1), with five of them added in the 1990s. Those include tibial muscular dystrophy (OMIM 253600), a form of infantile cerebelloptic atrophy (OMIM 260565), northern epilepsy (OMIM 600143), gonadal dysgenesis (OMIM 233300), and a lethal metabolic syndrome (OMIM 603358).

The delineation of these genetic disorders as new disease entities was made in all cases by observant clinicians who carefully excluded known causes and paid attention to the family histories of individual cases. These successes suggest that new diseases awaiting discovery are likely to also occur in other countries, especially when a subisolate structure has allowed the enrichment of recessive mutations. All but two of the disease genes have been mapped, and the gene and its common mutations are known for 22 diseases. The successes in positional cloning have been recently reviewed in more detail (6, 68, 69).

Some unusual incidences characterized the positional cloning of different disease genes, often progressing on neighboring benches or in neighboring institutes. As shown in Table 1, some genes for completely unrelated diseases mapped very close to each other (e.g., *CSTB* and *AIRE*; *MUL* and *MKS1*), promoting joint efforts in physical mapping. In another coincidence, genes for very phenotypically different disorders turned out to belong to the same gene family (*SLC26A2*, mutated in diastrophic dysplasia, and *SLC26A3*, mutated in congenital chloride diarrhea). Finally, for more than one disease, the mutated gene happened to carry

**TABLE 1** Diseases commonly listed for Finnish disease heritage\*

<b>Disease</b>	<b>Gene, position</b>	<b>Main mutation (&gt;75%)</b>	<b>OMIM #</b>
Autosomal recessive, incidence >1:10,000			
Congenital nephrosis, Finnish type	NPHN, 19q31.1	2-BP DEL, 121CT	256300, 602716
Autosomal recessive, incidence 1:10,000–1:40,000			
Aspartyl-glucosaminuria	AGA, 4q32-q33	CYS163SER	208400
Infantile neuronal ceroid lipofuscinosis; Santavuori-Haltia disease	PPT1, 1p32	ARG122TRP	256730, 600722
Meckel syndrome, type 1	MKS1, 17q22-q23	?	249000
Hydrolethalus syndrome	11q23-q25	?	236680
Diastrophic dysplasia	SLC26A2, 5q32-33.1	-26, T-C, +2 (GT-to-GC transition in a splice donor site)	222600
Cartilage-hair hypoplasia	CHH, 9p13	?	250250
Ovarian dysgenesis; XX gonadal dysgenesis	FSHR, 2p21-p16	ALA189VAL	233300, 136435
Myoclonic epilepsy of Unverricht and Lundborg	CSTB, 21q22.3	12-mer expansion in promoter	254800, 601145
Lethal congenital contracture syndrome; Herva disease	LCCS, 9q34	?	253310
Autoimmune polyendocrinopathy syndrome, type I	AIRE, 21q22.3	ARG257TER	240300
Salla disease; sialuria, Finnish type	SLC17A5, 6q14-q15	ARG39CYS	604369, 604322
Congenital chloride diarrhea	SLC26A3, 7q22-q31.1	VAL317DEL	214700, 126650
Autosomal recessive, incidence <1:40000			
Mulibrey nanism	MUL, 17q22-q23	5-BP DEL, NT493-497	253250, 605073
Nonketotic hyperglycinemia	GLDC, 9p22	SER564ILE	238300
Peho syndrome	?	?	260565
Ornithine aminotransferase deficiency	OAT, 10q26	ARG180THR	258870
Lysinuric protein intolerance	SLC7A7, 14q11.2	1181, A-T, -2 (10-bp deletion beginning at 1181)	222700, 603593
Usher syndrome, type III	USH3, 3q21-q25	?	276902
Cohen syndrome	COH1, 8q22-q23	?	216550

TABLE 1 (Continued)

Disease	Gene, position	Main mutation (>75%)	OMIM #
Cornea plana type 2	KERA, 12q21.3-q22	ASN247SER	217300, 603288
Infantile-onset spinocerebellar ataxia	IOSCA, 10q24	?	271245
Tibial muscular dystrophy	TMD, 2q11	?	600334
Hereditary fructose intolerance	ALDOB, 9q22.3	Multiple	229600
Imerslund-Grasbeck syndrome; megaloblastic anemia type 1	CUBN, 10p12.1	PRO1297LEU	261100, 602997
Northern epilepsy; progressive epilepsy with mental retardation	CLN8, 8pter- p22	ARG24GLY	600143
Finnish lethal neonatal metabolic syndrome	FLNMS, 2q33-q37	?	603358
Muscle-eye-brain disease	MEB, 1p34- p32	?	253280
Polycystic lipomembranous osteodysplasia with sclerosing Leuko- encephalopathy; Nasu- Hakola syndrome	TYROBP, 19q31.1	EX1-4 DEL	221770, 604142
Neuronal ceroid lipofuscinosis, late infantile type, Finnish variant	CLN5, 13q21.1-q32	2467AT DEL, TER	256731
Lactase deficiency	2q21	?	223000
Rapadilino syndrome	?	?	266280
Autosomal dominant			
Finnish type amyloidosis; amyloidosis V	GSN, 9q34	ASP187ASN	105120, 137350
X-linked recessive			
Choroideremia	CHM, Xq21.2	Multiple	303100
X-linked juvenile retinoschisis 1	RS1, Xp22.2- p22.1	Multiple	312700

\*The list includes 32 autosomal recessive diseases, 1 dominant, and 2 X chromosomal recessive. The autosomal recessive diseases are listed in approximate order of incidence; for the rarest diseases, the order is arbitrary. A single predominant founder mutation (>75% of chromosomes) is listed; for other entries, the mutation(s) remain unknown (?) or there are multiple mutations. For references on individual diseases, OMIM entry numbers are given (52).

an associated polymorphism (AGA, SLC26A3). Distinguishing the functional mutation from the polymorphism was finally accomplished by functional assays.

On the other hand, random sampling and the subsequent drift have produced a Finnish disease heritage with unexpectedly low frequencies of some genes that are much more common elsewhere. The most prominent examples include phenylketonuria, of which only four cases have ever been diagnosed in Finland (10), and cystic fibrosis, which shows regional founder effects with different mutations (28).

## Other Recessive Diseases

All recessive diseases that occur in Finland, however, do not show as much remarkable homogeneity of mutations as many of the 35 Finnish diseases. There are already many diseases in Finland where multiple mutations have been identified. Some of these examples include diseases of the Finnish heritage: The first to be identified was ornithine aminotransferase deficiency (OMIM 258870) where the most prevalent Finnish mutation, Leu402Pro, accounted for only ~85% of all mutations. At the time of its finding, this news was surprising; but investigators quickly understood that the relative enrichment of one mutation would serve to pick up all additional mutations in the population (as affected compound heterozygotes), including those with frequencies so low that homozygotes would be extremely rare. Another example of a Finnish disease with multiple mutations is retinoschisis (OMIM 312700). It shows both southern and northern geographic clusters in Finland, and its relative overrepresentation is caused by three widespread founder mutations (16).

The study of recessive diseases that are not particularly overrepresented in Finland in comparison to other European populations has often revealed multiple mutations. Each of them, however, shows distinct geographic clustering that is indicative of founder effects. Examples of these diseases include transglutaminase-1 (TGM1) mutations in autosomal recessive congenital ichthyoses and mapping of a new ichthyose locus with a founder location (38, 87), sulfonylurea receptor (SUR1) mutations in persistent hyperinsulinemic hypoglycemia of infancy (64), and coagulation factor XIII mutations in a rare bleeding disorder (53).

Especially illustrative are the results on steroid 21-hydroxylase gene (CYP21) mutations (45, 46). Homozygous CYP21 deficiency causes congenital adrenal hyperplasia and virilization in girls. The gene maps within the class III gene cluster in the HLA region in chromosome 6p21 and has a nearby pseudogene (CYP21P). This genomic structure favors unequal crossover events and frequent deletions as a result; indeed, the majority of CYP21 mutations are genomic rearrangements. The severe salt-wasting and virilizing forms are not uncommon and occur at a frequency of about 1/15,000 births in different populations (66). The location of the CYP21 gene within the HLA cluster makes it a highly interesting marker for population studies as well because the extreme polymorphism of HLA haplotypes and their associations with CYP21 mutations can be utilized. Although some CYP21 mutation-HLA haplotype correlations are enriched in some populations, in mixed

populations CYP21 variants do not show significant linkage disequilibrium with HLA haplotypes. This is consistent with the high mutation rate and recurrent independent mutation events for CYP21. How does this rather mutation-prone gene behave in Finland?

Levo et al (45) studied the number of different mutations that were present in two thirds of all diagnosed congenital adrenal hyperplasia patients in Finland, representing 74 unrelated families. They identified a total of 19 different mutation-haplotype combinations, consistent with the expectedly high mutation rate. However, three mutations that occurred in otherwise very rare HLA haplotypes accounted for half of all mutations (46). Over 80% of all mutation-haplotype combinations were observed repeatedly, suggesting founder mutations. Indeed, plotted on a map of Finland based on grandparental birthplaces, each mutation-haplotype combination showed clustering consistent with founder effects. Some mutations were probably imported and old, others were more likely to have first arisen locally. Notably, a genealogical analysis revealed no consanguinity between the 74 families back to the grandparental level, but extended analysis allowed the construction of some multiply consanguineous pedigrees with common ancestors in the 1600s and 1700s (up to 11 generations back). For most families, common ancestors could not be identified, suggesting only remote links. These results again support the picture of Finland as a country of multiple subisolates. The process of isolate formation and their expansion has been so recent and rapid that it also predominates over the diversifying effect of new mutations for this relatively mutable gene.

## Dominant Diseases

For dominant genes, the founder effects are equally visible. One dominant disorder, amyloidosis V or Finnish-type amyloidosis, is listed among the 35 Finnish diseases, and it has a single founder mutation in the gelsolin gene (Table 1). In long QT syndrome, a familial cardiac arrhythmia syndrome, a founder mutation was recently identified in Finland (72). Among more common dominant diseases, acute intermittent porphyria is illustrative. It is caused by mutations in the porphobilinogen deaminase gene. Forty known families in Finland possess a total of 26 mutations; the most common of them form founder clusters when plotted on a map (26, 54, 55).

Distinct founder effects have been observed in familial hypercholesterolemia (FH), inherited nonpolyposis colon cancer, and breast cancer. One specific mutation in the low-density lipoprotein lipase receptor gene (LDLR), called FH-North Karelia, accounts for almost 90% of cases in an eastern subpopulation of about 180,000, with at least 340 carriers identified (89). In all of Finland, four mutations (FH-North Karelia, FH-Helsinki, FH-Turku, and FH-Pori) account for about three quarters of all patients; FH-North Karelia alone is responsible for two thirds. The distribution of mutations varies in different parts of the country (30, 31). Interestingly, these mutations are not common for other Nordic countries: A survey among Swedish FH patients revealed only 5.5% (10 cases) of FH-Helsinki and a single

case of FH-North Karelia (48). This observation is consistent with the expected consequences of rapid recent population growth in a subisolate structure and little admixture with neighbors.

Two founder mutations in the *MLH1* gene are responsible for about two thirds of hereditary nonpolyposis colorectal cancer families in Finland (60–62), and mutations in the *BRCA1* and *BRCA2* genes have given rise to multiple founder effects in inherited breast cancer (76).

## LESSONS LEARNED

### Utility of Linkage Disequilibrium

What have we learned from the study of rare recessive diseases? Perhaps the most notable general lessons have been associated with the repeated successes in making use of the genetic founder effects, also called linkage disequilibrium mapping. This is based on the expectation that a single mutation has been increased in frequency by founder effect and drift for most recessive diseases that have a nonuniform geographic distribution. Introduced originally in a single copy, the allelic composition of that chromosome has eroded over time by recombinations, but it is still preserved near the disease gene. When a chromosome that carries a disease gene (found homozygous in affected individuals) is studied with polymorphic microsatellite markers, certain alleles for each marker show remarkable overrepresentation when compared to the general allele frequencies in the population (usually calculated on the basis of the untransmitted parental chromosomes). The alleles form conserved haplotypes that are repeatedly observed in patients from seemingly unrelated families. Linkage disequilibrium in this context refers to the association of certain alleles and haplotypes to the presence of the disease gene, and its significance is assessed by contingency tables and  $\chi$ -square tests comparing marker allele frequencies in patient chromosomes versus control chromosomes. One then looks for markers that are monomorphic in patient chromosomes or show the strongest allelic association: They are most likely to be closest to the disease gene.

A common misconception is that, given the relatively young age of the population subisolates, the conserved chromosomal segments are too long to provide sufficient resolution for positional cloning. This has, however, not been a practical problem. Although it is true that the conserved segments often span several cM and marker alleles of up to 5 cM away may show statistically significant associations, the conserved segments do not fully overlap between different families. The chromosomes that carry disease genes have undergone different historical recombinations in varying branches of the superpedigree that describe the descent of the common mutation. Thus, by compiling the haplotypes and comparing the likely positions of historical recombinations, one can infer the position of the disease gene very accurately.

This is exemplified in Figure 4, which shows haplotype data for congenital chloride diarrhea (14). Compilation of the haplotypes from only 24 core families (most of them with just one affected child) mapped the gene exactly between two markers, even though marker allele associations were statistically significant as far as 5–6 cM away from the disease gene.

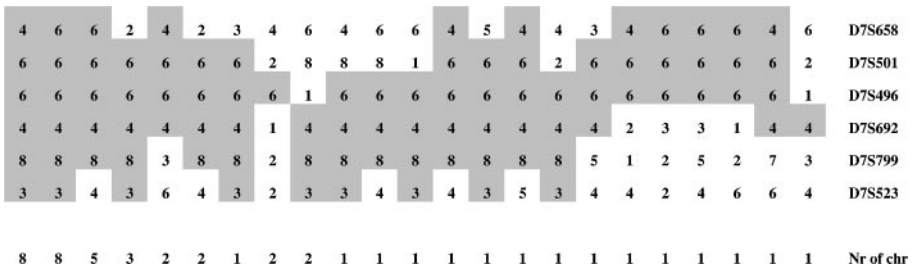
Further accuracy for the predictions of gene localizations was obtained from the application of Luria & Delbrück’s equation, originally introduced to assess mutations in bacterial cultures (49). The reinvention of the equation for the use of estimating gene position in isolated population settings was made by E. Lander, working with J. Hästbacka & A. de la Chapelle on the positional cloning of the diastrophic dysplasia gene (11, 12). The formula was simplified (43) to

$$p_{\text{excess}} = \alpha(1 - \Theta)^g \approx \alpha e^{-\Theta g}, \text{ where } p_{\text{excess}} = (p_{\text{affected}} - p_{\text{control}})/(1 - p_{\text{control}}).$$

In this equation,  $\alpha$  denotes the fraction of chromosomes that are expected to have a common mutation;  $g$ , the number of generations;  $\Theta$ , recombination distance from marker to mutation; and  $p_{\text{affected}}$  and  $p_{\text{control}}$ , the overrepresented allele frequency for the marker in disease gene-carrying and control chromosomes, respectively.

The accuracy and limits of the estimates became the subject of statistical debate (22, 23). Later, de la Chapelle & Wright (5) considered sources of error for the estimates and introduced a corrected standard error estimate for the Luria-Delbrück method.

In spite of confidence limits that were too narrow and the need to guess the number of generations since the founding of the branching mutation pedigree,



**Figure 4** Six marker haplotypes in chromosomes from patients with congenital chloride diarrhea [adapted from (14)]. The microsatellite markers, spanning 12 cM, are indicated on the right. Each haplotype observed in patients is depicted as a vertical column of six alleles, with the number of times it was encountered (bottom row). In 16 chromosomes, the haplotype was 4 or 6-6-6-4-8-3, suggesting the founder haplotype with an early recombination at D7S658 (indicated as 4/6 variation for that marker). The completely conserved haplotype segments in the remaining chromosomes are highlighted with gray. All chromosome segments overlap only between D7S496 and D7S692, suggesting a likely position for the gene. The gene is located 290 kb from D7S496 toward D7S692 (13).

application of the Luria-Delbrück estimations proved highly useful for practical positional cloning purposes. Refined mapping of the disease gene by the haplotype compilation or Luria-Delbrück methods was presented as an interlude to cloning for most of the diseases listed in Table 1. The endgame of positional cloning was played in some instances by isolating entirely new genes; but in later cases, the presence of positional candidate genes facilitated direct mutation analyses.

After genes became identified, the Luria-Delbrück equation was also used to estimate the age of mutation since its introduction to the expanding population, even though the statistical accuracy of this estimate might well be questioned. From the equation above,  $g$  was solved for several markers based on  $p_{\text{excess}}$  in mutation-carrying chromosomes and the distances from markers to the mutation, and the estimate was obtained by averaging overmarker-specific  $g$ 's (13). For example, the founder mutation in congenital chloride diarrhea was estimated to have spread for about 19 generations, in good accordance with the age of the eastern Finnish expansion (400 years).

## Study Design and Sampling

A population structure with small breeding units, large local variations in allele frequencies, and distinct subpopulations should make one cautious about extrapolations and predictions based on simple population models and analytical calculations (91). The unorthodox population structure should be already considered when applying the Hardy-Weinberg equation for estimating carrier frequencies from disease incidence. Also, predictions of levels of linkage disequilibrium based on a global view of Finland as a panmictic population (33) yield values that are too small in comparison to those observed for many rare genes. More complex models are clearly needed to aid in designing studies. In ongoing work, we are attempting to integrate pedigree and meiosis simulation with geographical frameworks to yield population models that might more accurately reflect the genetic structure of a population (V. Ollikainen, H. Mannila & J. Kere, unpublished).

In genetic studies, sampling is important, all the more so in a population that has a distinct substructure caused by founder effects and genetic drift, as is the case in rural Finland. In disease gene studies, cases and controls need to be genetically matched. One of the best ways to achieve this is to use family-based controls (84). If such a strategy cannot be adopted, then at least the origin of cases and controls should be controlled and matched geographically, considering potential pitfalls of population substructure.

These considerations are even more important when common diseases become subjects for genetic study. Relatively large variations for disease frequencies have been observed in epidemiological studies among different parts of the country, and these differences have been attributed in part to genetic variances (20, 81). Closer analysis shows that clinical and laboratory parameters also seem to profile patients differently in varying parts of the country, which further emphasizes the need to pay careful attention to sampling (4, 44, 80).

Especially when large numbers of coding SNPs become available, studies that remain inconclusive are all too easy to plan and execute. A tempting outline for a study might be, for example, the following: One could start by looking for geographic differences in the frequency of a common disease and sample families from relative high-frequency and low-frequency areas. For the next step, one could look at differences in the frequencies of candidate gene SNPs in both sample sets. If one would find a certain SNP at a significantly higher frequency in the high disease frequency area compared to the low frequency area, that might be interpreted as providing useful information about genes that contribute to the disease. This conclusion, however, would most likely be wrong. As shown already for both rare and common neutral alleles (e.g., the blood group markers), significant differences in their frequencies exist between various communities within Finland. Such differences affect all genes and are caused purely by the random effects of the low numbers of founders and the drift in small breeding units. Thus, almost any gene may have variances in frequency within two communities if they are small enough, and these variances may even be significant for a large number of genes. The phenomenon of different disease frequency in two communities may well be caused by founder events and drift, but one cannot tell by simple association which of the enriched genes contributes to the disease pathogenesis and which was enriched coincidentally.

## FOCUS ON MULTIFACTORIAL DISEASES

### First Experiences from Linkage Approaches

First experiences from mostly linkage-based projects aimed at identifying susceptibility genes in common diseases have started to accumulate from the Finnish population. The spectrum of diseases is already wide, including type 2 diabetes (9, 50, 90), multiple sclerosis (34), asthma (38a, 39), myocardial infarction (67), familial combined hyperlipidemia (65), hypertension (70), obesity (63), schizophrenia (8, 15), psoriasis (1), and systemic lupus (32). Several other projects are in progress. It is too early to draw conclusions based on this experience, but some first observations are possible.

The geographic clustering noted for some phenotypes has obviously not resulted from the simple enrichment of a single mutation. Three notable examples include type 2 diabetes, with high incidence on west coastal regions (50); multiple sclerosis, which is also enriched in a western subpopulation (34); and schizophrenia (8, 15). In schizophrenia, one genome scan was performed in the young subpopulation of Kuusamo (a high-incidence region), and another on sibpairs from other parts of Finland. Both studies suggest the presence of multiple loci (8, 15), and convincing haplotype association was not found. On the other hand, a genome scan on hypertension identified one locus as the most significant in Finland (70), and a genome-wide scan for asthma susceptibility genes in the Kainuu subpopulation resulted in significant linkage mapping of one locus, suggesting higher

homogeneity than in general populations (38a). As long as individual susceptibility genes and their disease-associated polymorphisms remain unidentified, not much more can be said about the success of these lines of work.

A strong interest in haplotype association as a tool to extract more information from the isolated population settings has sparked the development of statistical and computing methods (17, 41, 51, 85). At least in simulated settings, these methods have shown promise (77).

## Linkage Disequilibrium and Common Markers

The prospect of utilizing linkage disequilibrium to map common disease genes has sparked much recent interest. Several studies have recently addressed the extent of linkage disequilibrium in the population of Finland as well as elsewhere (7, 19, 71, 82, 86). The results have uniformly indicated that levels of linkage disequilibrium between nearby markers in Finland do not differ from levels observed in other European populations. Some authors have interpreted this to suggest that for the mapping of common disease genes studies on populations, such as those in Finland, might offer no specific advances.

The results are not unexpected, but the interpretation may be too generalized (2, 17, 82). Studies on overall levels of linkage disequilibrium have used samples collected from healthy individuals. As genetic markers, the studies have employed either microsatellites or SNPs, and they have especially considered the most common alleles for both types of markers. The common alleles have obviously been imported in the form of thousands of copies with the first and later settlers to Finland, and they even have been spread to the subisolates in tens or hundreds of copies, depending on each allele's frequency, and undergone recombinations at rates common to all populations. Thus, the overall measures of linkage disequilibrium for common alleles should be similar in Finland and elsewhere, especially when sampling is made on the "general" Finnish population.

Taillon-Miller et al (82) considered linkage disequilibrium for common SNP alleles in the Kainuu subpopulation. The levels of linkage disequilibrium were indistinguishable from, for example, the levels observed in CEPH samples. These data are consistent with the fact that Kainuu was founded by more than a few individuals. Interestingly, the haplotype patterns for two tight X-chromosomal clusters of markers were remarkably different between Kainuu and Sardinia. This result illustrates simply that the two isolates have different population histories and that different haplotypes have been traded differently by drift. Overall, these data are consistent with the previous finding that in local isolates in Finland even common blood group alleles may show significant deviations from the population mean. Similar results were obtained for the Kuusamo subisolate (86). Some linkage disequilibrium was observed between X-chromosomal markers in Kuusamo but very little between autosomal markers in both Kuusamo and a general Finnish sample set. Do these results invalidate linkage disequilibrium mapping for susceptibility genes in common multifactorial diseases?

## Linkage Disequilibrium in Disease Gene Studies

Whenever patients are collected, the sample set will become enriched for disease or susceptibility alleles at the same time as the sample represents a random sampling of alleles that are irrelevant for disease pathogenesis (or at least that is what investigators hope to achieve). It is important to note that this is true even for SNPs within one gene: A disease gene may contain both disease-causing SNPs and neutral SNPs, and depending on the history of each polymorphism, the latter may or may not be associated to the disease. Thus, a random SNP in a susceptibility gene may be totally useless for linkage disequilibrium mapping, which should be considered when designing SNP association studies. For example, a polymorphic gene (HCR) in the psoriasis susceptibility region PSORS1 contained SNPs that were tightly associated with psoriasis, as expected, but also a large number of presumably neutral unassociated SNPs (1).

The level of enrichment for any disease allele is dependent on both the biology and pathogenesis of the disease (how well the disease status predicts a defect in a particular gene) and on the genetic makeup of the population from which the samples were collected (how many different disease-causing alleles there are to choose from). The significance of disequilibrium observed around a susceptibility gene will then depend on the frequencies of the surrounding alleles and haplotypes in patient and control chromosomes.

In an extreme case, a disease is caused by a single mutation that occurs homozygous in all patients. That extreme case is what has been observed for many recessive monogenic diseases in Finland, and such single mutations are likely to be in strong linkage disequilibrium with the surrounding alleles and haplotype. On the other hand, any fully penetrant dominant gene is likely to be found in only half of all patient chromosomes, and linkage disequilibrium will thus be smaller when markers on all chromosomes are considered. With decreasing levels of penetrance, increasing proportion of phenocopies, and increasing levels of allelic heterogeneity behind the disease in the particular population that was sampled, linkage disequilibrium will diminish, ultimately becoming undetectable. If there is no enrichment for any particular allele, as is the case when normal individuals are sampled, one will observe only "background" disequilibrium. But by selecting for disease (i.e., introducing an intended bias), the expected level of linkage disequilibrium in the sample cannot be predicted from background disequilibrium.

## Power Considerations

It may be useful to consider issues of power when linkage disequilibrium analyses are used as part of the mapping strategy for susceptibility genes in common diseases. For example, Laitinen et al (39) recruited asthma patients and their families from the Kainuu region for susceptibility gene studies. The origin of the subjects was verified from population registries, and several candidate gene regions were considered (25, 39, 40). Haplotype association analysis

was employed in addition to genetic linkage analysis to increase power, and the investigators estimated that the study material would have allowed a good chance to identify an associated haplotype present in 15% of all patient chromosomes. Indeed, moderate but statistically significant haplotype associations were observed for two candidate gene regions, around the FCER2 gene in chromosome 19p and around the IL9R gene in the Xq pseudoautosomal region (25, 40).

How should such associations be interpreted? Of course, a false positive association remains a possibility. On the other hand, the population history of Kainuu may be one that is made up of founders who brought along a small number of susceptibility alleles for these genes that are neither necessary nor sufficient to induce the development of asthma in their carriers. Drift modified the allele frequencies so that perhaps one of them became more common than the others. When present-day patients were sampled, those specific susceptibility alleles with their surrounding haplotypes became further enriched by the process of sampling so that the difference between patient and control chromosomes became detectable. Recent considerations on the nature and frequency of disease-causing alleles that may well be relatively rare in general populations (93) would be compatible with this hypothesis. Clearly, until functional differences that are specific for associated alleles and compatible with disease pathogenesis have been demonstrated, the results only suggest a path for further study.

## A MODEL POPULATION: PROMISES AND LIMITATIONS

The allelic diversity of a disease in a particular population is a most important issue when susceptibility gene studies are planned. As is evident from marker data, allelic diversity in any founder population that has grown in isolation is different from other populations, and the differences will be most prominent for relatively rare alleles. The varying frequencies of alleles in small subpopulations that are caused by drift and simplified allelic diversity might aid genetic studies when SNP association becomes a major approach (3). For example, the finding of a corneodesmosin gene allele that was previously suggested as a psoriasis susceptibility gene at a very high frequency in Kainuu helped to exclude it as a susceptibility allele (1).

The rarer an allele is, the more likely it is to get lost because of drift, but occasional rare alleles survive at unusually high frequencies (Figure 3). Many disease susceptibility alleles may turn out to have low general population frequencies and high allelic diversity (93). For a disease gene with a large number of different mutations, this may mean the loss of most mutated alleles but the preservation of one or a few major mutations in the population isolate, accompanied with associated haplotypes. Unfortunately, predicting how individual alleles have fared is not possible.

On the other hand, if the disease alleles are very common and confer only a slight increase in disease risk (have a low penetrance), their identification in an isolated population will become difficult. The main limiting problem is likely

to be practical: The population size in an isolated population may not be large enough to allow for the collection of a sufficient number of patients that is needed to show a small increase in susceptibility. This practical issue also becomes a problem with some suggested solutions to the dilemma of identifying susceptibility genes. The Saami population is characterized genetically as one without signs of expansion, but rather long-time isolation with constant population size; and such a situation might help to map some old and common mutations (35, 83). However, with less than 2000 individuals announcing Saami as their mother tongue (79), the population appears too small for the study of even the most common multifactorial diseases.

During the past decade, most of the work that was needed to dissect the molecular genetics of the Finnish disease heritage has been accomplished. This decade is devoted more to the study of multifactorial diseases in Finland as well as elsewhere. Although the outcome from these studies is still difficult to predict, positive signs have started to emerge (38a). Regardless, we should expect to learn much more about how different populations yield information that can be applied to the annotation of the human morbidity map.

#### ACKNOWLEDGMENTS

I thank Vesa Ollikainen for the computer simulation presented in this article. Members of our research group and staff of the Finnish Genome Center have contributed via numerous discussions that were essential for the contents of this review. Our work is supported by the Academy of Finland, Sigrid Jusélius Foundation, Finnish Technology Fund Tekes, Ulla Hjelt Fund, and Helsinki University Hospital research funds. JK is a member of Biocentrum Helsinki and the Center of Excellence for Disease Genetics, University of Helsinki.

**Visit the Annual Reviews home page at [www.AnualReviews.org](http://www.AnualReviews.org)**

#### LITERATURE CITED

1. Asumalahti K, Laitinen T, Itkonen-Vatjus R, Lokki ML, Suomela S, et al. 2000. A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-associated susceptibility allele. *Hum. Mol. Genet.* 9:1533–42
2. Boehnke M. 2000. A look at linkage disequilibrium. *Nat. Genet.* 25:246–47
3. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22:231–38
4. Carmelli D, Williams RR, Rissanen A. 1982. Contrasting patterns of familiarity for cholesterol and triglyceride in Finland according to type of coronary manifestations and locations. *Am. J. Epidemiol.* 116:617–21
5. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton Univ. Press
6. de la Chapelle A, Wright FA. 1998. Linkage disequilibrium mapping in isolated populations: the example of Finland

- revisited. *Proc. Natl. Acad. Sci. USA* 95:12416–23
7. Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, et al. 2000. The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 25:320–23
  8. Ekelund J, Lichtermann D, Hovatta I, Elilonen P, Suvisaari J, et al. 2000. Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22. *Hum. Mol. Genet.* 9:1049–57
  9. Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, et al. 2000. The Finland–United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.* 67:1174–85
  10. Guldberg P, Henriksen KF, Sipilä I, Gutler F, de la Chapelle A. 1995. Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. *J. Med. Genet.* 32:976–78
  11. 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.* 2:204–11
  12. Hästbacka J, de la Chapelle A, Mah-tani MM, Clines G, Reeve-Daly MP, et al. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 78:1073–87
  13. Höglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, et al. 1996. Mutations in the down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. *Nat. Genet.* 14:316–19
  14. Höglund P, Sistonen P, Norio R, Holmberg C, Dimberg A, et al. 1995. Fine mapping of the congenital chloride diarrhoea gene by linkage disequilibrium. *Am. J. Hum. Genet.* 57:95–102
  15. Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, et al. 1999. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. *Am. J. Hum. Genet.* 65:1114–24
  16. Huopaniemi L, Rantala A, Forsius H, Somer M, de la Chapelle A, Alitalo T. 1999. Three widespread founder mutations contribute to high incidence of X-linked juvenile retinoschisis in Finland. *Eur. J. Hum. Genet.* 7:368–76
  17. Jorde LB. 2000. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* 10:1435–44
  18. Jorde LB, Pitkänen KJ. 1991. Inbreeding in Finland. *Am. J. Phys. Anthropol.* 84:127–39
  19. Jorde LB, Watkins WS, Kere J, Nyman D, Eriksson AW. 2000. Gene mapping in isolated populations: new roles for old friends? *Hum. Hered.* 50:57–65
  20. Jousilahti P, Vartiainen E, Tuomilehto J, Pekkanen J, Puska P. 1998. Role of known risk factors in explaining the difference in the risk of coronary heart disease between eastern and southwestern Finland. *Ann. Med.* 30:481–87
  21. Jutikkala E. 1959. *Suomen Historian Kartasto (Atlas of Finnish History)*. Porvoo: Finnish Acad. Sci. 2nd ed.
  22. Kaplan NL, Hill WG, Weir BS. 1995. Likelihood methods for locating disease genes in nonequilibrium populations. *Am. J. Hum. Genet.* 56:18–32
  23. Kaplan NL, Weir BS. 1995. Are moment bounds on the recombination fraction between a marker and a disease locus too good to be true? Allelic association mapping revisited for simple genetic diseases in the Finnish population. *Am. J. Hum. Genet.* 57:1486–98
  24. Karjalainen E. 1989. Migration and regional development in the rural communes of Kainuu, Finland in 1980–1985. *Nordia* 23:1–89

25. Kauppi P, Laitinen T, Ollikainen V, Mannila H, Laitinen LA, Kere J. 2000. The IL9R region contribution in asthma is supported by genetic association in an isolated population. *Eur. J. Hum. Genet.* 8:788–92
26. Kauppinen R, Mustajoki S, Pihlaja H, Peltonen L, Mustajoki P. 1995. Acute intermittent porphyria in Finland: 19 mutations in the porphobilinogen deaminase gene. *Hum. Mol. Genet.* 4:215–22
27. Keränen J. 1984. Kainuun asuttaminen (the settling of Kainuu). *Studia Historica Jyväskyläensia*, Vol. 28. Jyväskylä: Univ. Jyväskylä. 282 pp.
28. Kere J, Estivill X, Chillón M, Morral N, Nunes V, et al. 1994. Cystic fibrosis in a low-incidence population: two major mutations in Finland. *Hum. Genet.* 93:162–66
29. Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, et al. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. *Am. J. Phys. Anthropol.* 108:381–99
30. Koivisto UM, Viikari J S, Kontula K. 1995. Molecular characterization of minor gene rearrangements in Finnish patients with heterozygous familial hypercholesterolemia: identification of two common missense mutations (Gly823-to-Asp and Leu380-to-His) and eight rare mutations of the LDL receptor gene. *Am. J. Hum. Genet.* 57:789–97
31. Kontula K, Koivisto UM, Koivisto P, Turtoola H. 1992. Molecular genetics of familial hypercholesterolaemia: common and rare mutations of the low density lipoprotein receptor gene. *Ann. Med.* 24:363–67
32. Koskenmies S, Widén E, Kere J, Julkunen H. 2001. Familial systemic lupus erythematosus in Finland. *J. Rheumatol.* 28:758–60
33. Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22:139–44
34. Kuokkanen S, Gschwend M, Rioux JD, Daly MJ, Terwilliger JD, et al. 1997. Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am. J. Hum. Genet.* 61:1379–87
35. Laan M, Pääbo S. 1997. Demographic history and linkage disequilibrium in human populations. *Nat. Genet.* 17:435–38
36. Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, et al. 1996. The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am. J. Hum. Genet.* 58:1309–22
37. Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, et al. 1999. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* 7:447–58
38. Laiho E, Ignatius J, Mikkola H, Yee VC, Teller DC, et al. 1997. Transglutaminase 1 mutations in autosomal recessive congenital ichthyosis: private and recurrent mutations in an isolated population. *Am. J. Hum. Genet.* 61:529–38
- 38a. Laitinen T, Daly MJ, Rioux JD, Kauppi P, Laprise C, et al. 2001. A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat. Genet.* 28:87–91
39. Laitinen T, Kauppi P, Ignatius J, Ruotsalainen T, Daly MJ, et al. 1997. Genetic control of serum IgE levels and asthma: linkage and linkage disequilibrium studies in an isolated population. *Hum. Mol. Genet.* 6:2069–76
40. Laitinen T, Ollikainen V, Lazaro C, Kauppi P, de Cid R, et al. 2000. Association study of the chromosomal region containing the FCER2 gene suggests it has a regulatory role in atopic disorders. *Am. J. Respir. Crit. Care Med.* 161:700–6
41. Lake SL, Blacker D, Laird NM. 2000. Family-based tests of association in the presence of linkage. *Am. J. Hum. Genet.* 67:1515–25
42. Lander ES, Schork NJ. 1994. Genetic

- dissection of complex traits. *Science* 265:2037–48
43. Lehesjoki AE, Koskiniemi M, Norio R, Tirrito S, Sistonen P, et al. 1993. Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum. Mol. Genet.* 2:1229–34
  44. Lehtinen S, Luoma P, Nayha S, Hassi J, Ehnholm C, et al. 1998. Apolipoprotein A-IV polymorphism in Saami and Finns: frequency and effect on serum lipid levels. *Ann. Med.* 30:218–23
  45. Levo A, Jääskeläinen J, Sistonen P, Sirén MK, Voutilainen R, Partanen J. 1999. Tracing past population migrations: genealogy of steroid 21-hydroxylase (CYP21) gene mutations in Finland. *Eur. J. Hum. Genet.* 7:188–96
  46. Levo A, Partanen J. 1997. Mutation-haplotype analysis of steroid 21-hydroxylase (CYP21) deficiency in Finland. Implications for the population history of defective alleles. *Hum. Genet.* 99:488–97
  47. Libert F, Cochaux P, Beckman G, Samson M, Aksenova M, et al. 1998. The delta-ccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in north-eastern Europe. *Hum. Mol. Genet.* 7:399–406
  48. Lind S, Eriksson M, Rystedt E, Wiklund O, Angelin B, Eggertsen G. 1998. Low frequency of the common Norwegian and Finnish LDL-receptor mutations in Swedish patients with familial hypercholesterolaemia. *J. Intern. Med.* 244:19–25
  49. Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511
  50. Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, et al. 1996. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat. Genet.* 14:90–94
  51. McIntyre LM, Martin ER, Simonsen KL, Kaplan NL. 2000. Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. *Genet. Epidemiol.* 19:18–29
  52. McKusick VA. 1998. *Mendelian Inheritance in Man*. Baltimore: Johns Hopkins Univ. Press. 12th ed. <http://www.ncbi.nlm.nih.gov/omim>
  53. Mikkola H, Syrjälä M, Rasi V, Vahtera E, Hämäläinen E, et al. 1994. Deficiency in the A-subunit of coagulation factor XIII: two novel point mutations demonstrate different effects on transcript levels. *Blood* 84:517–25
  54. Mustajoki P, Desnick RJ. 1985. Genetic heterogeneity in acute intermittent porphyria: characterisation and frequency of porphobilinogen deaminase mutations in Finland. *Br. Med. J.* 291:505–9
  55. Mustajoki S, Pihlaja H, Ahola H, Petersen NE, Mustajoki P, Kauppinen R. 1998. Three splicing defects, an insertion, and two missense mutations responsible for acute intermittent porphyria. *Hum. Genet.* 102:541–48
  56. Nevanlinna HR. 1972. The Finnish population structure. A genetic and genealogical study. *Hereditas* 71:195–236
  57. Nevanlinna HR. 1980. Rare hereditary diseases and markers in Finland: an introduction. In *Population Structure and Genetic Disorders*, ed. A Eriksson. London: Academic
  58. Norio R, Nevanlinna HR, Perheentupa J. 1973. Hereditary diseases in Finland; rare flora in rare soil. *Ann. Clin. Res.* 5:109–41
  59. Nunez MG. 1987. A model for the early settlement of Finland. *Fennosc. Archaeol.* 4:3–18
  60. Nyström-Lahti M, Kristo P, Nicolaides NC, Chang SY, Aaltonen LA, et al. 1995. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat. Med.* 1:1203–6
  61. Nyström-Lahti M, Sistonen P, Mecklin JP, Pylkkänen L, Aaltonen LA, 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* 91:6054–58
62. Nyström-Lahti M, Wu Y, Moisio AL, Hofstra RM, Osinga J, et al. 1996. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. *Hum. Mol. Genet.* 5:763–69
63. Ohman M, Oksanen L, Kaprio J, Koskenvuo M, Mustajoki P, et al. 2000. Genome-wide scan of obesity in Finnish sibpairs reveals linkage to chromosome Xq24. *J. Clin. Endocrinol. Metab.* 85:3183–90
64. Otonkoski T, Ämmälä C, Huopio H, Cote GJ, Chapman J, et al. 1999. A point mutation inactivating the sulfonylurea receptor causes the severe form of persistent hyperinsulinemic hypoglycemia of infancy in Finland. *Diabetes* 48:408–15
65. Pajukanta P, Terwilliger JD, Perola M, Hiekkalinna T, Nuotio I, et al. 1999. Genomewide scan for familial combined hyperlipidemia genes in Finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am. J. Hum. Genet.* 64:1453–63
66. Pang SY, Wallace MA, Hofman L, Thuline HC, Dorche C, et al. 1988. World-wide experience in newborn screening for classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Pediatrics* 81:866–74
67. Pastinen T, Perola M, Niini P, Terwilliger J, Salomaa V, et al. 1998. Array-based multiplex analysis of candidate genes reveals two independent and additive genetic risk factors for myocardial infarction in the Finnish population. *Hum. Mol. Genet.* 7:1453–62
68. Peltonen L. 2000. Positional cloning of disease genes: advantages of genetic isolates. *Hum. Hered.* 50:66–75
69. Peltonen L, Jalanko A, Varilo T. 1999. Molecular genetics of the Finnish disease heritage. *Hum. Mol. Genet.* 8:1913–23
70. Perola M, Kainulainen K, Pajukanta P, Terwilliger JD, Hiekkalinna T, et al. 2000. Genome-wide scan of predisposing loci for increased diastolic blood pressure in Finnish sibs. *J. Hypertens.* 18:1579–85
71. Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A, Slatkin M, Freimer NB. 1995. The distribution of linkage disequilibrium over anonymous genome regions. *Hum. Mol. Genet.* 4:887–94
72. Piippo K, Laitinen P, Swan H, Toivonen L, Viitasalo M, et al. 2000. Homozygosity for a HERG potassium channel mutation causes a severe form of long QT syndrome: identification of an apparent founder mutation in the Finns. *J. Am. Coll. Cardiol.* 35:1919–25
73. Pitkänen K, Jorde LB, Mielke JH, Fellman JO, Eriksson AW. 1988. Marital migration and genetic structure in Kitee, Finland. *Ann. Hum. Biol.* 15:23–33
74. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, et al. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* 5:42–52
75. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S. 1996. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl. Acad. Sci. USA* 93:12035–39
76. Sarantaus L, Huusko P, Eerola H, Launonen V, Vehmanen P, et al. 2000. Multiple founder effects and geographical clustering of BRCA1 and BRCA2 families in Finland. *Eur. J. Hum. Genet.* 8:757–63
77. Sevon P, Ollikainen V, Onkamo P, Toivonen HTT, Mannila H, Kere J. 2001. Mining associations between genetic markers, phenotypes and covariates. *Genet. Epidemiol. (Suppl.)* In press
78. Siren MK, Sareneva H, Lokki ML, Koskimies S. 1996. Unique HLA antigen frequencies in the Finnish population. *Tissue Antigens* 48:703–7

79. Statistics Finland. 1997. *Statistical Yearbook of Finland*, Vol. 92. Hämeenlinna: Statistics Finland
80. Stengård JH, Kardia SL, Tervahauta M, Ehnholm C, Nissinen A, Sing CF. 1999. Utility of the predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men aged 65 to 84 years is dependent on context defined by Apo E genotype and area of residence. *Clin. Genet.* 56:367–77
81. Suhonen O, Reunanen A, Aromaa A, Knekt P, Pyörälä K. 1985. Four-year incidence of myocardial infarction and sudden coronary death in twelve Finnish population cohorts. *Acta Med. Scand.* 217:457–64
82. Taillon-Miller P, Bauer-Sardiña I, Saccone NL, Putzel J, Laitinen T, et al. 2000. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat. Genet.* 25:324–28
83. Terwilliger JD, Zollner S, Laan M, Pääbo S. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: “drift mapping” in small populations with no demographic expansion. *Hum. Hered.* 48:138–54
84. Thomson G. 1995. Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* 57:487–98
85. Toivonen HTT, Onkamo P, Vasko K, Ollikainen V, Sevón P, et al. 2000. Data mining applied to linkage disequilibrium mapping. *Am. J. Hum. Genet.* 67:133–45
86. Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L. 2000. Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. *Eur. J. Hum. Genet.* 8:604–12
87. Virolainen E, Wessman M, Hovatta I, Niemi K-M, Ignatius J, et al. 2000. Assignment of a novel locus for autosomal recessive congenital ichthyosis to chromosome 19p13.1–p13.2. *Am. J. Hum. Genet.* 66:1132–37
88. Virtaranta-Knowles K, Sistonen P, Nevanlinna HR. 1991. A population genetic study in Finland: comparison of the Finnish- and Swedish-speaking populations. *Hum. Hered.* 41:248–64
89. Vuorio AF, Turtola H, Piilahti KM, Repo P, Kanninen T, Kontula K. 1997. Familial hypercholesterolemia in the Finnish north Karelia. A molecular, clinical, and genealogical study. *Arterioscler. Thromb. Vasc. Biol.* 17:3127–38
90. Watanabe RM, Ghosh S, Langefeld CD, Valle TT, Hauser ER, et al. 2000. The Finland–United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am. J. Hum. Genet.* 67:1186–200
91. Workman PL, Mielke JH, Nevanlinna HR. 1976. The genetic structure of Finland. *Am. J. Phys. Anthropol.* 44:341–67
92. Wright AF, Carothers AD, Pirastu M. 1999. Population choice in mapping genes for complex diseases. *Nat. Genet.* 23:397–404
93. Zwick ME, Cutler DJ, Chakravarti A. 2000. Patterns of genetic variation in Mendelian and complex traits. *Annu. Rev. Genomics Hum. Genet.* 1:387–407