Am. J. Hum. Genet. 56:654–662, 1995

# Batten Disease Gene, *CLN3*: Linkage Disequilibrium Mapping in the Finnish Population, and Analysis of European Haplotypes

Hannah M. Mitchison,[1] Angela M. O'Rawe,[1] Peter E. M. Taschner,[2] Lodewijk A. Sandkuijl,[2,3,4] Pirkko Santavuori,[5] Nanneke de Vos,[2] Martijn H. Breuning,[2] Sara E. Mole,[1] R. Mark Gardiner,[1] and Irma E. Järvelä[1]

[1]Department of Paediatrics, University College London Medical School, London; [2]Department of Human Genetics, Leiden University, Leiden; [3]Institute of Clinical Genetics, Erasmus University, Rotterdam; [4]Department of Human Genetics, University of Groningen, Groningen; and [5]Department of Child Neurology, University of Helsinki, Helsinki

## Summary

The gene for Batten disease (juvenile-onset neuronal ceroid lipofuscinosis, or Spielmeyer-Sjögren disease), *CLN3*, maps to 16p11.2-12.1. Four microsatellite markers—*D16S288, D16S299, D16S298,* and *SPN*—are in strong linkage disequilibrium with *CLN3* in 142 families from 16 different countries. These markers span a candidate region of ~2.1 cM. *CLN3* is most prevalent in northern European populations and is especially enriched in the isolated Finnish population, with an incidence of 1:21,000. Linkage disequilibrium mapping was applied to further refine the localization of *CLN3* in 27 Finnish families by using linkage disequilibrium data and information about the population history of Finland to estimate the distance of the closest markers from *CLN3*. *CLN3* is predicted to lie 8.8 kb (range 6.3–13.8 kb) from *D16S298* and 165.4 kb (132.4–218.1 kb) from *D16S299*. Enrichment of allele "6" at *D16S298* (on 96% of Finnish and 92% of European *CLN3* chromosomes) provides strong evidence that the same major mutation is responsible for Batten disease in Finland as in most other European countries and that it is therefore not a Finnish mutation. Genealogical studies show that Batten disease is widespread throughout the densely populated regions of Finland. The ancestors of two Finnish patients carrying rare alleles "3" and "5" at *D16S298* in heterozygous form originate from the southwestern coast of Finland, and these probably represent other foreign mutations. Analysis of the number and distribution of *CLN3* haplotypes from 12 European countries provides evidence that more than one mutation has arisen in Europe.

## Introduction

Batten disease, or juvenile-onset neuronal ceroid lipofuscinosis (JNCL; also known as Spielmeyer-Sjögren disease) is the most common form of neuronal ceroid lipofuscinosis (NCL) in childhood (Santavuori 1988). It is an autosomal recessive, progressive encephalopathy characterized by visual failure, seizures, psychomotor deterioration, and accumulation of autofluorescent lipopigment resembling ceroid and lipofuscin in neural and nonneural tissues. The primary defect is unknown. The first symptoms appear at 4–7 years of age, and life expectancy is ~25 years. The key diagnostic criterion is vacuolated lymphocytes in peripheral blood. Typical fingerprint profiles of storage bodies are found in electron-microscopic investigations (Santavuori 1988). Published estimates of incidence vary between 0.71/100,000 and 4/100,000 (Santavuori 1988; Claussen et al. 1992). The exact number of Batten patients is known in Finland: 170 cases have been diagnosed by one of the authors (P.S.) since 1969. The disease affects 1:21,000 live births, with a carrier frequency of 1/70, which indicates a specific enrichment in this population. The gene for Batten disease, *CLN3*, has been assigned to 16p11.2-12.1, and four highly polymorphic microsatellite markers, *D16S288, D16S299, D16S298,* and *SPN*, are in strong linkage disequilibrium with the disease gene in 142 affected families from 16 different countries. This suggests a founder effect in the disease (Mitchison et al. 1993, 1994; Lerner et al. 1994). The genetic distance between the flanking markers *D16S288* and *D16S383* is 2.1 cM (sex averaged). The haplotype "7561" formed by the linkage disequilibrium markers is present on 28% of *CLN3* chromosomes, and the core haplotype "56" (at markers *D16S299* and *D16S298*) is even more strongly enriched, present on 73% of *CLN3* chromosomes (Mitchison et al. 1994).

Batten disease has been diagnosed worldwide, but it is most common in northern European populations. Strikingly, in Finland JNCL is as prevalent as the most common autosomal recessive diseases that are specifically enriched in the Finnish population (Norio et al. 1973; Norio 1981). Therefore, its behavior in this isolated population is directly comparable to that of these diseases. Linkage disequilibrium mapping has been applied in studies of six diseases enriched in the Finnish population (Hästbacka et al. 1992; Lehesjoki et al. 1993; Aaltonen et al. 1994; Haataja et al. 1994; Kestilä et al. 1994; Savukoski et al. 1994) to achieve

Received June 23, 1994; accepted for publication December 6, 1994.

Address for correspondence and reprints: Dr. Irma Järvelä, Department of Paediatrics, University College London Medical School, 5 University Street, London WC1E 6JJ, United Kingdom.



0002-9297/95/5603-0014$02.00

a significant reduction in the estimated size of the region predicted to contain the disease gene. These efforts used the Luria-Delbrück equations (Luria and Delbrück 1943), originally developed for bacterial populations, which apply information from linkage disequilibrium and the population history to estimate the probable distance between the disease gene and the enriched allele of the closest marker loci. In the case of the recently cloned diastrophic dysplasia gene (DTDST), the calculation correctly predicted the position of the gene (Hästbacka et al. 1994). Linkage disequilibrium mapping is based on the assumption that, in the immediate vicinity of the gene, a distinctive haplotype is detected that reflects the ancestral chromosome carrying the mutation, introduced into the population sufficiently long ago that recombination has made the region of strongest linkage disequilibrium sufficiently small. The calculation requires knowledge about both the number of original settlers and the timing of settlement in an isolated population. The Luria-Delbrück formula has been applied using Finland as a model population because studies on the Finnish gene pool (Hästbacka et al. 1992) indicate that the Finnish population has expanded from a small number of settlers to the current population of 5 million during the past 2,000 years and has remained isolated because of the geographic and cultural barriers (Nevanlinna 1972; Norio et al. 1973; Norio 1981). As a consequence >30 diseases, mostly autosomal recessive, are enriched in the Finnish population that are very rare or lacking in other populations (Norio et al. 1973; Norio 1981).

Here we have applied linkage disequilibrium mapping to Finnish Batten families, as an isolated population, to further refine the predicted region that contains *CLN3*. Genealogical data were collected to verify the distribution of *CLN3* in the Finnish population and to locate *CLN3* haplotypes on the map of Finland. In direct comparison with the Finnish data, the number and distribution of *CLN3* haplotypes and alleles in other Scandinavian and European populations were analyzed.

## Families and Methods

### *Families*

Forty-three Finnish Batten families participated in the genealogical study. This represents 88% (43/49) of the total number of Batten families in Finland that have living children. The birthplaces of the great-grandparents were traced by interviewing the parents. Twenty-seven Finnish, 32 Dutch, 9 German, 11 Norwegian, 10 Danish, 9 British, 5 Swedish, 3 Italian, 2 Belgian, 1 Austrian, 1 Greek, and 1 Icelandic family were included in haplotype analysis. Four Finnish families had three affected children, 5 families had two affected children, and 18 families had one affected child. Altogether, 111 Batten families with 162 affected individuals were analyzed.

### *Linkage Disequilibrium and Haplotype Analysis*

Allele frequencies of four markers—*D16S288, D16S299, D16S298,* and *SPN,* as described by Mitchison et al. (1994)—were calculated for the Finnish family material to detect possible enrichment. Linkage disequilibrium was calculated using the $\chi^2$ test and Fisher's exact test using the Bonferroni correction. Haplotypes based on the two markers—*D16S299* and *D16S298*—known to form the "core" haplotype "56" (Mitchison et al. 1994) and for *D16S298* alone were analyzed in all European families.

### *Linkage-Disequilibrium Mapping*

In view of both the strength of linkage disequilibrium and the assumption that one major mutation causes Batten disease in Finland, Luria-Delbrück analysis (Luria and Delbrück 1943; Hästbacka et al. 1992) was used to calculate both allelic homogeneity at the four markers in strongest linkage disequilibrium with *CLN3* and the genetic distance from each of them to *CLN3*. The model is based on the following assumptions: (1) Finland was founded $g$ (=100) generations ago by a small initial population with $n_o$ chromosomes (=1,000), $n_o/2$ individuals, one of which carried the *CLN3* mutation; (2) the population increased at a constant rate of $e^d$-fold per generation, where $d$ = 0.0921 ($n = n_o e^{gd}$), to the current 5 million ($n$ = 10 million chromosomes) (Hästbacka et al. 1992); (3) for the known gene frequency, $q$ (=.007, under Hardy-Weinberg equilibrium), for *CLN3,* the number of *CLN3* chromosomes is $0.007n$; (4) the population has been panmictic, i.e., the *CLN3* chromosomes detected represent a random selection of those in the population.

The contribution of new mutations to the pool of Batten disease–causing chromosomes was calculated via $\alpha = 1 - m/q$, where $\alpha$ is the proportion of disease-causing chromosomes that descended from a single common ancestor. $m$ is the likely number of *CLN3* chromosomes carrying new mutations and is calculated according to the equations given by Hästbacka et al. (1992): $m = (\mu/d)\ln(10^7\mu/d) \pm 2(\mu/d)$, with $10^7$ representing the current number of chromosomes 16 in Finland and with $\mu$ representing the mutation rate. For an autosomal recessive disease it is not possible to calculate the mutation rate directly, so an average mutation rate, $\mu = 5 \times 10^{-6}$/generation for autosomal dominant and X-linked diseases, was used (Vogel 1990). To calculate the approximate distance of *CLN3* from the closest markers, it is assumed that the majority of disease-causing chromosomes in the Finnish population descended from a single ancestral chromosome that carried a particular allele, A1, at a nearby marker locus. Disease-causing chromosomes (1) that do not descend from the common ancestor but instead carry different mutations that either have arisen since the founding of Finland or have been contributed by immigrants or (2) in which recombination has taken place between the marker and the disease locus must be taken into account. The above chromosomes need

not carry A1, although they may do so by chance. Given this information, if $\alpha$ denotes the proportion of disease-causing chromosomes descended from the common ancestor, $\pi$ the proportion of disease-causing chromosomes descended from the common ancestor that have not undergone a recombination between *CLN3* and the marker locus, $p_{normal}$ the allele frequency of A1 in the general population, and $p_{affected}$ the significantly higher allele frequency of A1 on disease-bearing chromosomes, then $p_{affected} = \alpha\pi + (1 - \alpha\pi)p_{normal}$ (Lehesjoki et al. 1993). Given the observed frequencies of $p_{normal}$ and $p_{affected}$, the proportion, $\alpha\pi$, of *CLN3* chromosomes can be estimated as $\alpha\pi = p_{excess} = (p_{affected} - p_{normal})/(1 - p_{normal})$. $\pi$ can be seen as a function of the recombination fraction, $\theta$, between *CLN3* and the marker. Given $\pi_{obs} = p_{excess}/\alpha$, the corresponding value of $\theta$ can be calculated iteratively via the equations given by Hästbacka et al. (1992): $\pi_{obs} = (\theta/d)\ln(70{,}000\theta/d) \pm 2(\theta/d)$, where 70,000 indicates the estimated number of Batten disease–causing chromosomes in the current Finnish population.

## Results

### *Distribution of* CLN3 *in the Finnish Population*

The distribution of birthplaces of the great-grandparents of 43 Finnish Batten patients (fig. 1) shows that *CLN3* is widely spread through the most densely populated areas of Finland. It is prevalent on the southwestern coast, as well as in the middle and eastern parts of the country.

### *Linkage-Disequilibrium Analysis in the Finnish Population*

Alleles at loci *D16S288, D16S299, D16S298,* and *SPN* were analyzed in the Finnish families. The results are shown in table 1. All four markers are in strong linkage disequilibrium, as expected. The $\chi^2$ value at *D16S298* was the highest. No significant difference in the strength of linkage disequilibrium could be determined among the four loci. Allele "6" at *D16S298* is present in 96% of Finnish *CLN3* chromosomes and in 30% of normal chromosomes. In the other three markers the same tendency of enrichment was detected: *D16S299* allele "5" in 87% versus 24%, *D16S288* allele "7" in 75% versus 7%, and *SPN* allele "1" in 82% versus 36%, of *CLN3* chromosomes versus normal chromosomes, respectively. There was a smaller collection of both disease alleles and normal alleles in the Finnish chromosomes, compared with the other European populations analyzed here (table 2).

### *Estimating Genetic Distances by Linkage Disequilibrium*

The approximate genetic distances between microsatellite marker loci and *CLN3* were calculated using the formulas described in Families and Methods. By Luria-Delbrück analysis, $1 - \alpha$, the proportion of *CLN3* chromosomes carrying a mutation other than the founder mutation, is .04883 (likely range .033323–.064344), i.e., 4.9% (3.3%–6.4%). Therefore 95.1% (93.6%–96.7%) of the CLN3 chromosomes carry the founder mutation ($\alpha$ = .951167; likely range .935656–.966677). Three $\alpha$ values covering the above range were used to estimate the distances to the markers in linkage disequilibrium with *CLN3*. To consider the maximum possible distances that could occur in this analysis, $\theta$ values were also calculated under the assumption of 100% homogeneity, i.e., $\alpha$ = 1.0 (table 3). Given these values of $\alpha$, the closest marker to *CLN3* is *D16S298,* which has a $\theta$ value of 0.00000–0.000762 (total range 0.0000–0.001038) with respect to the *CLN3* gene. For the next closest marker, *D16S299,* $\theta$ is calculated to be 0.001489–0.002128 (total range 0.001189–0.00278). The corresponding values for *D16S288* and *SPN* are 0.002646–0.003173 (0.002137–0.00409) and 0.002839–0.003355 (0.002302–0.004317) respectively. *D16S383,* the proximal flanking marker, has not been analyzed, because it has very low informativeness and negligible linkage disequilibrium in *CLN3* chromosomes.

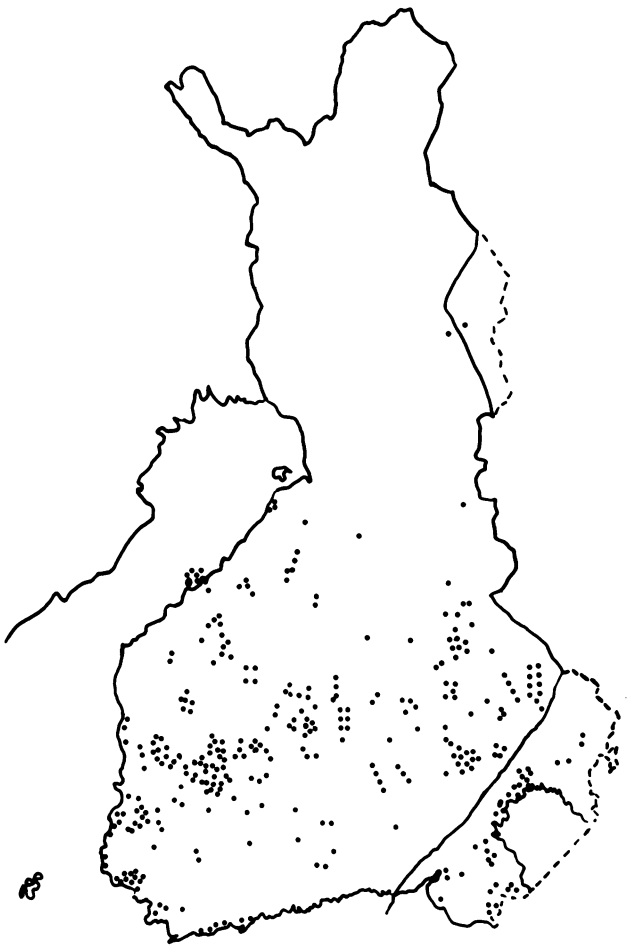

**Figure 1** Birthplaces of great-grandparents of 43 *CLN3* patients in Finland. Families with multiple affected children are presented only once.

## Table 1

**Chromosome 16 Dinucleotide Repeat Loci in Linkage Disequilibrium in Finnish Families**

| Locus and Allele Size (bp) | Allele Designation | No. (Frequency) CLN3 Chromosomes | No. (Frequency) Non-CLN3 Chromosomes | $\chi^2$ (1 df) | *P* |
|---|---|---|---|---|---|
| D16S288: | | | | | |
| 166 ..... | 1 | 0 (.00) | 0 (.00) | | |
| 164 ..... | 2 | 2 (.045) | 0 (.00) | | |
| 162 ..... | 3 | 1 (.02) | 18 (.41) | | |
| 160 ..... | 4 | 3 (.07) | 8 (.18) | | |
| 158 ..... | 5 | 2 (.045) | 6 (.14) | | |
| 156 ..... | 6 | 3 (.07) | 9 (.20) | | |
| 142 ..... | 7 | 33 (.75) | 3 (.07) | | |
| Null ..... | 8 | 0 (.00) | 0 (.00) | | |
| 154 ..... | 9 | 0 (.00) | 0 (.00) | | |
| 152 ..... | 10 | 0 (.00) | 0 (.00) | | |
| 146 ..... | 11 | 0 (.00) | 0 (.00) | | |
| Overall ..... | | 44 (1.00) | 44 (1.00) | 42.31 | <.001 |
| D16S299: | | | | | |
| 126 ..... | 1 | 0 (.00) | 2 (.04) | | |
| 124 ..... | 2 | 0 (.00) | 1 (.02) | | |
| 122 ..... | 3 | 0 (.00) | 4 (.07) | | |
| 120 ..... | 4 | 3 (.055) | 12 (.22) | | |
| 118 ..... | 5 | 47 (.87) | 13 (.24) | | |
| 116 ..... | 6 | 3 (.055) | 18 (.33) | | |
| 114 ..... | 7 | 1 (.02) | 2 (.04) | | |
| 112 ..... | 8 | 0 (.00) | 1 (.02) | | |
| 110 ..... | 9 | 0 (.00) | 1 (.02) | | |
| Overall ..... | | 54 (1.00) | 54 (1.00) | 43.35 | <.001 |
| D16S298: | | | | | |
| 190 ..... | 1 | 0 (.00) | 1 (.02) | | |
| 188 ..... | 2 | 0 (.00) | 4 (.07) | | |
| 186 ..... | 3 | 1 (.02) | 19 (.35) | | |
| 184 ..... | 4 | 0 (.00) | 9 (.17) | | |
| 182 ..... | 5 | 1 (.02) | 5 (.09) | | |
| 180 ..... | 6 | 52 (.96) | 16 (.30) | | |
| 178 ..... | 7 | 0 (.00) | 0 (.00) | | |
| Null ..... | 8 | 0 (.00) | 0 (.00) | | |
| 170 ..... | 9 | 0 (.00) | 0 (.00) | | |
| Overall ..... | | 54 (1.00) | 54 (1.00) | 51.46 | <.001 |
| SPN: | | | | | |
| 251 ..... | 1 | 36 (.82) | 16 (.36) | | |
| 252 ..... | 2 | 3 (.07) | 4 (.09) | | |
| 253 ..... | 3 | 3 (.07) | 5 (.12) | | |
| 254 ..... | 4 | 1 (.02) | 1 (.02) | | |
| 255 ..... | 5 | 0 (.00) | 4 (.09) | | |
| 256 ..... | 6 | 1 (.02) | 4 (.09) | | |
| 257 ..... | 7 | 0 (.00) | 2 (.05) | | |
| 258 ..... | 8 | 0 (.00) | 3 (.07) | | |
| 259 ..... | 9 | 0 (.00) | 3 (.07) | | |
| 260 ..... | 10 | 0 (.00) | 1 (.02) | | |
| 261 ..... | 11 | 0 (.00) | 0 (.00) | | |
| 262 ..... | 12 | 0 (.00) | 0 (.00) | | |
| 263 ..... | 13 | 0 (.00) | 0 (.00) | | |
| 264 ..... | 14 | 0 (.00) | 1 (.02) | | |
| 268 ..... | 15 | 0 (.00) | 0 (.00) | | |
| 249 ..... | 16 | 0 (.00) | 0 (.00) | | |
| 250 ..... | 17 | 0 (.00) | 0 (.00) | | |
| 317 ..... | 18 | 0 (.00) | 0 (.00) | | |
| Overall ..... | | 44 (1.00) | 44 (1.00) | 18.81 | <.001 |

## Table 2

**Distribution of CLN3 Haplotypes in European Populations**

| | Haplotype | Finland | Sweden | Norway | Denmark | Iceland | United Kingdom | The Netherlands | Belgium | Germany | Austria | Italy | Greece | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *56* | 47 | 9 | 20 | 15 | 1 | 14 | 50 | 3 | 12 | 2 | | 2 | 175 |
| 2 | *66* | 2 | 1 | 1 | | | | 5 | | | | 2 | | 11 |
| 3 | *36* | | | | | | | 3 | | 5 | | | | 8 |
| 4 | *46* | 2 | | | 2 | | 2 | 2 | | | | | | 8 |
| 5 | *76* | 1 | | | | | | | | | | | | 1 |
| 6 | *26* | — | — | — | 1 | — | — | — | 1 | — | — | — | — | 2 |
| Overall for subgroup | **6* | 52 | 10 | 21 | 18 | 1 | 16 | 60 | 4 | 17 | 2 | 2 | 2 | 205 |
| 7 | *73* | | | | | 1 | | | | | | | | 1 |
| 8 | *63* | 1 | | | | | 1 | | | | | 2 | | 4 |
| 9 | *53* | | | 1 | | | | | | | | | | 1 |
| 10 | *43* | — | | — | 1 | — | — | | | | | — | | 1 |
| Overall for subgroup | **3* | 1 | | 1 | 1 | 1 | 1 | | | | | 2 | | 7 |
| 11 | *44* | | | | | | | 2 | | | | 1 | | 3 |
| 12 | *64* | | | | | | 1 | | | | | | | 1 |
| 13 | *54* | | | | | | — | 1 | | | | — | | 1 |
| Overall for subgroup | **4* | | | | | | 1 | 3 | | | | 1 | | 5 |
| 14 | *55* | | | | | | | | | 1 | | | | 1 |
| 15 | *65* | | | | 1 | | | | | | | | | 1 |
| 16 | *45* | 1 | | | — | | | | | — | | | | 1 |
| Overall for subgroup | **5* | 1 | | | 1 | | | | | 1 | | | | 3 |
| 17 | *47* | — | — | — | — | — | — | — | 1 | — | — | — | 1 | 2 |
| Total | | 54 | 10 | 22 | 20 | 2 | 18 | 64 | 4 | 18 | 2 | 6 | 2 | 222 |

**Table 3**

**Estimation of Genetic Distances between *CLN3* and the Four Marker Loci in Linkage Disequilibrium, for Various Values of α**

| Locus | $p_{excess}$ | θ (Range) | | | |
|---|---|---|---|---|---|
| | | α = .935656 | α = .951167 | α = .966677 | α = 1.00 |
| *D16S288* ..... | .731707 | .002646 (.002137–.003424) | .0027776 (.00225–.003594) | .002907 (.002359–.003358) | .003173 (.002579–.00409) |
| *D16S299* ..... | .829268 | .001489 (.001189–.001931) | .001654 (.001324–.002181) | .001811 (.001453–.00238) | .002128 (.001714–.00278) |
| *D16S298* ..... | .947368 | .00000 (.0000) | .000088 (.000063–.000138) | .000333 (.000253–.000473) | .000762 (.000597–.001038) |
| *SPN* ..... | .714285 | .002839 (.002302–.003672) | .002970 (.002411–.003837) | .00309 (.002516–.003996) | .003355 (.00273–.004317) |

### *Analysis of European Haplotypes*

The number and distribution of core haplotypes represented by two of the closest markers, *D16S299* and *D16S298,* were analyzed in 12 different European countries. The results are shown in table 2. The core haplotype "56" is the most common haplotype (175/222 [79%]) in all European countries. The small numbers of families (and therefore chromosomes) analyzed from certain countries allow limited conclusions about their dominant haplotype to be drawn. The frequency of the "56" haplotype on disease-causing chromosomes is higher in Scandinavian countries compared with other European populations: 87% (47/54) of Finnish, 91% (20/22) of Norwegian, and 90% (9/10) of Swedish *CLN3* chromosomes. The frequency in the Danish population (75% [15/20]) closely resembles those of the other European populations: Dutch 78% (50/64), British 78% (14/18), and German 67% (12/18). The three Italian families had uncommon "core" haplotypes "66/66," "63/63," and "44/47," in contrast to "56/56." When the frequencies of allele "6" from the closest marker locus *D16S298* among disease-causing chromosomes are compared, the enrichment is still more striking: 96% (52/54) of Finnish, 100% (10/10) of Swedish, 95% (21/22) of Norwegian, 94% (60/64 and 17/18) of Dutch and German, 90% (18/20) of Danish, and 89% (16/18) of British *CLN3* chromosomes carry allele "6" at *D16S298.*

The two Finnish patients with either allele "3" or allele "5," instead of allele "6," at *D16S298* have ancestors from the southwestern coast. Allele "3" is also present in most Nordic countries, as well as in the United Kingdom and Italy. Allele "5" at *D16S298* is also present in Denmark and Germany. These alleles may therefore represent two other rare mutations brought to Finland from abroad.

## Discussion

Isolated populations have been a valuable resource in linkage disequilibrium studies for refining the region of interest in many inherited diseases such as familial Mediterranean fever (Aksentijevich et al. 1993), hemochromatosis (Yaouanq et al. 1994), and childhood-onset spinal muscular atrophy (Simard et al. 1994). However, the results have not always been unambiguous (Sirugo et al. 1993), and traditional linkage disequilibrium does not give precise information about distances between the gene of interest and the markers (Hill and Weir 1994).

In young, growing populations linkage disequilibrium mapping can be used to estimate the distance between a gene and closely linked markers, as well as to estimate the age of the mutation in the population, if the θ value is known (Luria and Delbrück 1943; Hästbacka et al. 1992). The accuracy of these calculations has been confirmed for DTDST (Hästbacka et al. 1994), in which the predicted distance of 64 kb from the nearest marker has proved to be very close to the actual value of 70 kb. Retrospective analysis for aspartylglucosaminuria (AGU), an autosomal recessive lysosomal storage disease strongly enriched in the Finnish population, gives some evidence supporting the usefulness of this method. More than 200 AGU patients have been described from Finland, which has a local carrier frequency of 1/30–1/80, whereas only ~40 patients are known from other parts of the world (Hietala et al. 1993). The spectrum of mutations causing AGU in Finland as well as in other populations has been characterized recently. Ninety-eight percent of Finnish patients carry the same $AGU_{Fin}$ point mutation ($Cys_{163}$→Ser), and this is not found either among patients of non-Finnish origin (Ikonen et al. 1991) or among >600 normal blood samples from Finno-Ugrian populations so far analyzed (P. Aula, personal communication). This suggests that the $AGU_{Fin}$ mutation has arisen among the Finns. To test the model that is the basis for the calculations in this paper, we have applied some of the formulas to AGU. If the distribution of mutations causing AGU were not known, then, by the formula $\alpha = 1 - m/q$ as described in Families and Methods, where $q = .017$, $\mu = 5 \times 10^{-6}$, and $d = 0.0921$, α is calculated to be .98 at a carrier frequency of 1/30 and .92 at a carrier frequency of 1/80. The observed value for the whole of

Finland is .98. Similarly, the age of the mutation (i.e., $g$) can be calculated by using the related formula $\alpha = 1 - \mu g/q$ (Hästbacka et al. 1992). $\alpha = .98$; $q = .017$, and $\mu = 5 \times 10^{-6}$/generation, which gives $g$ a value of 68 generations. This means that the age of $AGU_{Fin}$ mutation is approximated to be 1,360–1,700 years (20–25 years/generation). This calculation is consistent with the age of the $AGU_{Fin}$ mutation being less than the age of the founding population and therefore supports the model used in our calculations.

Batten disease serves as an example of an inherited disease that is present worldwide but also is enriched in an isolated population, Finland. Strong linkage disequilibrium was detected both in mixed populations with four markers spanning 2.1 cM (Mitchison et al. 1993, 1994; Lerner et al. 1994) and, by the present study, in the isolated Finnish population. The smaller number of disease and normal alleles present in the Finnish Batten disease population is consistent with the status of an isolated population and reflects the smaller number of chromosomes brought by the founders and available for recombination events. With marker *D16S298,* allele "6" was detected in 92% (205/222) of European disease-bearing chromosomes and, separately, in 96% (52/54) of Finnish disease-bearing chromosomes, compared with only 30% of normal chromosomes, supporting the assumption that the same major mutation is enriched in both Europe and Finland. Even without detailed mathematical analysis, one can assume that *CLN3* may lie very close to *D16S298.*

The genealogical study demonstrates that the birthplaces of the great-grandparents are widely spread throughout the earliest and most densely populated area of Finland. This is consistent with the Finnish population model used here, which assumes that a reasonably long time was needed for the population carrying *CLN3* to spread over such a wide area of the country. However, it is not possible to differentiate between the *CLN3* haplotype contribution by the founding population and that contributed by later immigrants, because both populations would probably have brought the same major haplotype and the same mutation into Finland. On the basis of the allelic association data and the Luria-Delbrück calculations presented here, it is reasonable to assume that ~95% of Finnish chromosomes carry the same mutation. If genetic distances using $\alpha = .951167$ are converted to physical distances, under the assumption that 1 cM is ~1 Mb, then *CLN3* will be positioned 8.8 kb (range 6.3–13.8 kb) from *D16S298,* 165.4 kb (132.4–218.1 kb) from *D16S299,* 277.7 kb (225.0–359.4 kb) from *D16S288,* and 297.0 kb (241.1–383.7 kb) from *SPN.* An extreme upper limit for the distances was estimated by assuming 100% homogeneity, which, to judge from the number and distribution of haplotypes, is clearly not the case. These are therefore the maximum distances possible under these assumptions. The predicted distances are compatible with both the predicted order of the genetic markers (Shen et al. 1994) and the known physical order of the markers (authors' unpublished results). The actual physical distances between markers *D16S299* and *D16S298* in this region of chromosome 16 are smaller than those predicted from the genetic map assuming 1 cM = 1 Mb, so, in this region of chromosome 16, it is more accurate to take 1 cM < 1 Mb (authors' unpublished results). Consequently, in this region the Luria-Delbrück formula would position *CLN3* even closer to *D16S299* and *D16S298.*

The accuracy of these calculations should be considered as informative but not conclusive, since the population model used here contains many assumptions and represents only one view of the complicated history of Finland. Estimates concerning the date of founding and the size of the initial population can only be based on available archaeological and historical data. However, reasonably wide variations in the time of founding ($g$) are acceptable in the Luria-Delbrück model, without any dramatic effect on the predicted genetic distance. The value of $g$ was assumed to be 100 in this study. For Batten disease, $g$ could be greater because the same major haplotype was found in all Scandinavian and Middle European populations, and so the founder mutation on the core haplotype may have already existed in the small ancestral population, before subsequent major settlement. If so, the predicted distances between the markers and *CLN3* would be even shorter. Later introduction of the disease-causing mutation into the Finnish population—e.g., only 40 generations, or 800–1,000 years ago—would approximately double the estimated distance between *D16S298* and *CLN3* (data not shown). This change is so small that it would not affect the cloning strategy for *CLN3.* Precise knowledge about the time of founding would be more important in those Finnish diseases in which the estimated distance between the marker and the disease is larger.

The factors that could significantly reduce $\alpha$ and therefore decrease the predicted distances are a higher mutation rate ($\mu$) or a higher proportion of immigrants who brought other mutations into the country. Two of 54 Finnish chromosomes have either allele "3" or allele "5," instead of allele "6," at D16S298. Allele "3" is also present in Norway, Denmark, Iceland, the United Kingdom, and Italy, whereas allele "5" is also present in Denmark and Germany. Therefore, it may be possible that no new mutations have arisen in Batten disease during the Finnish population history, in which case the distances calculated could be an overestimate. A greater $q$ value could increase the value of $\alpha$. Our estimate of the number of patients in Finland, which affects the value of $q$, is almost certainly correct, because of the long tradition of studying NCL in Finland. Therefore, 95% is a reasonable value of $\alpha$ to use in calculations of genetic distances by the Luria-Delbrück formula.

Mutations or slippage (Weber 1990) at the microsatellites (*D16S288, D16S299, D16S298,* and *SPN*) would also

affect the calculation, by decreasing the observed linkage disequilibrium. *D16S288* and *D16S299* showed no non-Mendelian alleles in 111 JNCL families (300 meioses) so far studied. However, at *D16S299,* alleles "4" and "6," which are 2 bp longer and shorter, respectively, than the most common allele, allele "5," may have arisen by slippage. At *D16S298* the only unusual alleles that have been observed are in two patients, one from The Netherlands (Moroccan family) and the other from Finland, where microsatellite mutations are probably not responsible for the situation. In the Moroccan patient, who is part of a consanguineous family, both alleles are deleted at this locus (P. E. M. Taschner and M. H. Breuning, unpublished data), and in the Finnish patient apparent non-Mendelian inheritance can be explained by the presence of a deletion on one paternal chromosome. These independent findings may in fact support our linkage disequilibrium mapping results defining *D16S298* as the marker closest to *CLN3,* since the suggestion, in both cases, is that a deletion across that locus exists that disrupts the *CLN3* gene. *SPN* is a complex dinucleotide repeat marker (Mitchison et al. 1994) that has twice given a non-Mendelian result. These families were excluded from the present study.

Seventy-nine percent of European *CLN3* chromosomes carry the core haplotype "56" at markers *D16S299-D16S298.* Allele "6" at *D16S298* is present even more often, on 92% of disease-causing chromosomes. Five other alleles at *D16S299* are combined with "6" at *D16S298* and probably represent ancient recombination events. Allele "4" and allele "7" at *D16S298* are not found in Scandinavian populations, whereas allele "3" and allele "5" are more scattered. Data from Mediterranean countries are too scarce to make further conclusions about other founder effects, but the observed haplotypes support the conclusion that more than one mutation exists in Europe. When the Batten disease data presented in this paper are compared with the data from cystic fibrosis (CF) (Morral et al. 1994), it is reasonable to assume that, among Europeans, a smaller number of mutations will be found in Batten disease than in CF.

In conclusion, the linkage disequilibrium mapping method applied, in this study, to the isolated population of Finland has defined the region containing the *CLN3* gene to be a maximum of 13.8 kb from *D16S298.* Theoretically this map resolution is at the level where only one YAC or cosmid is needed in order to identify the gene. It will be possible to evaluate the final value of this method when the actual *CLN3* gene has been identified. In the meantime, these results will focus cloning efforts to identify the CLN3 gene, by considerably narrowing the chromosome region that needs to be analyzed.

## Acknowledgments

We are grateful to the Finnish Batten disease families for providing their family data for this study. We thank James Elmslie, Atholl Johnson, Patricia Munroe, and Minna Savukoski for helpful discussions. The financial support of the Sigrid Jusélius Foundation (Finland), The Wellcome Trust and The Medical Research Council (United Kingdom), The Children's Brain Disease Foundation and the National Institutes of Health (United States) are acknowledged.

## References

Aaltonen J, Björses P, Sandkuijl LA, Perheentupa J, Peltonen L (1994) An autosomal locus causing autoimmune disease: autoimmune polyglandular disease type I assigned to chromosome 21. Nat Genet 8:83–87

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosen L, et al (1993) Familial Mediterranean fever (FMF) in Moroccan Jews: demonstration of a founder effect by extended haplotype analysis. Am J Hum Genet 53:644–651

Claussen M, Heim P, Knispel J, Goebel HH, Kohlschutter A (1992) Incidence of Neuronal ceroid-lipofuscinoses in West-Germany: variation of a method for studying autosomal recessive disorders. Am J Med Genet 42:536–538

Haataja L, Schleutker J, Laine A-P, Renlund M, Savontaus M-L, Dib C, Weissenbach J, et al (1994) The genetic locus for free sialic acid storage disease maps to the long arm of chromosome 6. Am J Hum Genet 54:1042–1049

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

Hästbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, et al (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine structure linkage disequilibrium mapping. Cell 78:1073–1087

Hietala M, Grön K, Syvänen A-C, Peltonen L, Aula P (1993) Prospects of carrier screening of aspartylglucosaminuria in Finland. Eur J Hum Genet 1:296–300

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet 54:705–714

Ikonen E, Aula P, Grön K, Tollersrud O, Halila R, Manninen T, Syvänen AC, et al (1991) Spectrum of mutations in aspartylglucosaminuria. Proc Natl Acad Sci USA 88:11222–11226

Kestilä M, Männikkö M, Holmberg C, Gyapay G, Weissenbach J, Savolainen E-R, Peltonen L, et al (1994) Congenital nephrotic syndrome of the Finnish type maps to the long arm of chromosome 19. Am J Hum Genet 54:757–764

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

Lerner TJ, Boustany R-M, MacCormack K, Gleitsman J, Schlumpf K, Breakefield XO, Gusella JF, et al (1994) Linkage disequilibrium between the juvenile neuronal ceroid lipofuscinosis gene and marker loci on chromosome 16p12.1. Am J Hum Genet 54:88–94

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

Mitchison HM, Taschner PEM, O'Rawe AM, de Vos N, Phillips HA, Thompson AD, Kozman HM, et al (1994) Genetic mapping of the Batten disease locus (*CLN3*) to the interval *D16S288*-D16S383 by analysis of haplotypes and allelic association. Genomics 22:465–468

Mitchison HM, Thompson AD, Mulley JC, Kozman HM, Richards RI, Callen DF, Stallings RL, et al (1993) Fine genetic mapping of Batten disease locus (*CLN3*) by haplotype analysis and demonstration of allelic association with chromosome 16p microsatellite loci. Genomics 16:455–460

Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Gimenez J, Reis A, et al (1994) The origin of the major cystic fibrosis mutation (Δ508) in European populations. Nat Genet 7:169–175

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

Norio R (1981) The diseases of Finland and Scandinavia. In: Rothschild H (ed) Biocultural aspects of disease. Academic Press, New York, pp 359–415

Norio R, Nevanlinna HR, Perheentupa J (1973) Hereditary diseases in Finland: rare flora in rare soil. Ann Clin Res 5:109–141

Santavuori P (1988) Neuronal ceroid lipofuscinoses in childhood. Brain Dev 10:80–83

Savukoski M, Kestilä M, Williams R, Järvelä I, Sharp J, Harris J, Santavuori P, et al (1994) Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. Am J Hum Genet 55:695–701

Shen Y, Kozman H, Thompson A, Phillips HA, Holman K, Nancarrow J, Lane S, et al (1994) A PCR-based genetic linkage map of human chromosome 16. Genomics 22:68–76

Simard LR, Prescott G, Rochette C, Morgan K, Lemieux B, Mathieu J, Melancon SB, et al (1994) Linkage disequilibrium analysis of childhood-onset spinal muscular atrophy (SMA) in the French-Canadian population. Hum Mol Genet 3:459–463

Sirugo G, Cocozza S, Brice A, Cavalcanti F, de Michele G, Dones I, Filla A, et al (1993) Linkage disequilibrium analysis of Freidreich's ataxia in 140 caucasian families: positioning of the disease locus and evaluation of allelic heterogeneity. Eur J Hum Genet 1:133–143

Vogel F (1990) Mutation in man. In: Emery AEH, Rimoin DL (eds) Principles and practice of medical genetics. Vol 1, 2d ed. Churchill Livingstone, New York, pp 56–57

Weber JL (1990) Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. In: Davies KE, Tilghman SM (eds) Genome analysis: genetic and physical mapping. Vol 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 159–181

Yaouanq J, Prichon M, Chorney M, Pontarotti P, Le Treut A, El Kahloun A, Mauvieux V, et al (1994) Anonymous marker loci within 400 kb of *HLA-A* generate haplotypes in linkage disequilibrium with the hemochromatosis gene (*HFE*). Am J Hum Genet 54:252–263