The gene for congenital chloride diarrhea maps close to but is distinct from the gene for cystic fibrosis transmembrane conductance regulator

**JUHA KERE*, PERTTI SISTONEN†, CHRISTER HOLMBERG‡, AND ALBERT DE LA CHAPELLE***

*Department of Medical Genetics and †Children's Hospital, University of Helsinki, 00014 Helsinki, Finland; and §Finnish Red Cross Blood Transfusion Service, 00310 Helsinki, Finland

Communicated by David Botstein, August 16, 1993

**ABSTRACT**

Congenital chloride diarrhea (CLD) is characterized by watery stools with high chloride content beginning prenatally and is inherited as an autosomal recessive trait. Perfusion studies have established a basic defect in ileal and colonic Cl−/HCO3− transport, resulting in defective chloride absorption. The protein and its gene defects have, however, remained uncharacterized. We attempted to exclude candidate genes by considering linkage disequilibrium as well as genetic linkage in a small number of Finnish families. Initial results were suggestive of linkage between CLD and the cystic fibrosis transmembrane regulator gene (CFTR). Extended analysis in eight families confirmed close linkage to chromosome 7 markers proximal of CFTR, with maximum logarithm of odds scores of 5.11 and 5.06 for D7S501 and D7S496, respectively, at zero recombination. Allelic associations were observed that were striking between CLD and D7S496 and weaker between CLD and D7S501. Multipoint analyses mapped CLD unequivocally at D7S496 with a maximum logarithm of odds score of 9.33. We conclude that the CLD gene maps close to, but is distinct from, CFTR.

Originally described in American patients of partly Italian descent (1, 2), congenital chloride diarrhea (CLD; MIM 214700, ref. 3) has been reported in close to a hundred patients in at least 14 countries in Europe, North America, and Asia (refs. 4-7 and the references therein). Approximately one-third of CLD patients live in Finland, but a high incidence has also been observed in Poland and Kuwait (4-7). CLD is inherited as an autosomal recessive trait and shows an uneven geographical distribution pattern within Finland (8); it clusters almost exclusively in a "late inhabited" region that was settled during the late 16th century.

The pathognomonic feature of CLD is watery diarrhea with a high chloride content (>90 mmol/liter when fluid and electrolyte deficiencies have been corrected). If suspected, the diagnosis is easy and can be readily established by stool electrolyte measurements (5). In vivo perfusion studies of the intestine have shown that the basic defect in CLD involves an anion-exchange mechanism in the ileum and colon. Normally chloride is transported in exchange for bicarbonate in both directions, resulting in net absorption from the lumen, but in CLD, both absorption and secretion are defective (9-11). Furthermore, animal models suggest that the Cl−/HCO3− exchange protein(s) are expressed in ileal brush border membrane, in both villus and crypt cells but are not expressed in basolateral membrane (12). Even though a Cl−/HCO3− exchange mechanism has been implicated in CLD, its molecular characteristics remain unknown.

As a first step toward identification of the CLD gene, we considered three candidate genes, hoping to implicate or exclude them by the combination of genetic linkage and linkage disequilibrium study. Encoding for Cl−/HCO3− exchange proteins, the erythrocyte band 3 surface protein gene (AE1; formerly EPB3) has been mapped to chromosome 17q12-q21 (13), and the highly homologous anion-exchanger 2 gene (AE2, formerly EPB3LI) has been mapped to 7q35-q36 (14). The cystic fibrosis transmembrane regulator gene (CFTR), mapping to chromosome 7q31-q32, is involved in chloride transport in sweat glands and mucosal membranes, and a role in bicarbonate trafficking has been suggested as well (15, 16). Restriction fragment length polymorphism (RFLP) markers for CFTR suggested the absence of recombinations between CLD and CFTR, prompting a more thorough study. We report here the mapping of the CLD gene to human chromosome 7q31, slightly proximal of CFTR, and a strong allelic association with the marker D7S496 in Finnish patients that is consistent with the hypothesis of a single founder.

**MATERIALS AND METHODS**

CLD Families. Eight Finnish families were studied; their structures are shown in Fig. 1. Blood samples for DNA extraction and linkage studies were collected from altogether 15 parents, 16 children affected with CLD, and 12 healthy siblings. The diagnosis was established within the first months of life in all patients who presented with hydramnios and watery stools with a high chloride content from birth. All patients have been treated with oral NaCl and KCl substitution from early infancy. Growth and development have been normal, and the diarrhea persists but does not cause limitations for the patients.

**Linkage Markers and Molecular Methods.** Twelve markers, eight of them highly informative dinucleotide repeats, were used (Table 1). Markers D7S58, D7S18, D7S23, and MET were analyzed for RFLPs as described (17); the following probe/restriction enzyme combinations were used: J3.11/Msp I, 7C22/EcoRI, KM.19/Pst I, CS.7/Hha I, XV-2c/Taq I, MetD/Taq I, and MetH/Msp I. CA-repeats for CFTR (17), D7S487, D7S490, D7S496, D7S501, D7S515, D7S523, and D7S530 (19) were studied by using PCR to amplify polymorphic genomic regions, followed by electrophoresis in 6% polyacrylamide gels. Detection of 32P-labeled PCR products was by autoradiography.

**Statistical Analyses.** Pairwise linkage analyses were done by using the programs MLINK and LODSCORE of the LINKAGE package (20). The programs CILINK and CDMAP were used for multipoint analyses. The data were haplotyped for RFLP markers MET and D7S23, consisting of two and three polymorphic systems, respectively.

For multipoint analyses the data were combined with that from the Centre d’Etude du Polymorphisme Humain (Paris) data base version 6.0 (1992) that included published data for all markers but CFTR and D7S23 (21). The highly significant

---

*Abbreviations: CLD, congenital chloride diarrhea; RFLP, restriction fragment length polymorphism; iod, logarithm of odds.*
assocation observed between D7S496 and CLD was applied to multipoint analyses by assuming no recombination in healthy siblings where haplotype analysis was ambiguous (altogether six meioses). This assumption introduced a risk that a possible recombination event was overlooked, but it was judged to be small, considering the small number of meioses and highly significant allelic association. The order and distances between markers were obtained by using the published order (19) for markers D7S497, D7S490, D7S496, D7S501, D7S515, D7S523, and D7S330 as a base order to compute odds for pairwise inversions between adjacent loci. MET and D7S38 markers were added consecutively in the best frame by calculating the odds of placement in each locus interval. Thereafter the odds for pairwise inversions were recalculated. CLD was placed by calculating the odds of its placement in each interval. All analyses were performed by using Kosambi mapping function, no sex difference, and complete penetrance for CLD.

The significance of allelic association between markers and CLD was calculated by Fisher’s exact test with Bonferroni correction and one-sided P values.

Haplotype Analysis. Haplotypes were constructed manually in each family by assuming the minimum number of recombinations.

RESULTS

CLD Is Linked to CFTR but Is Distinct from It. In preliminary analyses, the RFLP markers failed to show any recombinants with CLD. However, a different haplotype was observed in each CLD-carrying chromosome, contrary to the expectation were CFTR indeed the CLD gene (see Discussion). To allow conclusive analysis, highly informative markers were next used. Two-point linkage analysis disclosed no recombinations (Table 1), but multipoint and haplotype analyses revealed one recombinant (see below). No association between CFTR alleles and CLD was observed (Table 2).

Location of CLD and Order of Loci. The highly informative PCR markers D7S501, D7S496, D7S490, D7S330, D7S23, and D7S515 all showed conclusive evidence for linkage (Zmax ≥ 3). The highest two-point lod scores of 5.11 and 5.06 were obtained at zero recombination fraction with D7S501 and D7S496, respectively (Table 1). The closeness of these markers to CLD was further attested by the association of one allele (allele 6 for both markers) with CLD (Table 2). The association was much more pronounced for D7S496 than for D7S501 (Table 2).

Table 1. Pairwise logarithm of odds (lod) scores in eight Finnish CLD families with 12 markers from chromosome 7

Table 2. Allele distribution in eight unrelated Finnish CLD families

<table>
<thead>
<tr>
<th>Allele</th>
<th>D7S515</th>
<th>D7S501</th>
<th>D7S496</th>
<th>D7S523</th>
<th>CFTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>11*</td>
<td>2</td>
<td>14**</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values under each marker system denote the number of each allele observed in CLD and normal (N) chromosomes; zeros have been omitted for clarity. The distribution of alleles in normal chromosomes was derived from the unaffected chromosomes of the parents. *P = 0.01268; **P = 0.000002.

System refers to the method of detection and type of polymorphism (CA, dinucleotide repeat). The maximum lod scores (Zmax) and the respective recombination fractions (θ) are given in parentheses for computation results using the strong allelic association to D7S496 (see Results). Note the enhanced discrimination power.

*Approximate 90% confidence (conf.) limits were obtained by the ±1 lod unit method.
We considered this justified in view of the very strong allelic association. The risk is dependent on the actual recombination fraction between \( D7S496 \) and \( CLD \).

Haplotypes constructed assuming the minimum number of recombinations in each family disclosed one event that helped to map \( CLD \). In family 5, an unaffected daughter inherited the \( CLD \) haplotype from her father and the distal part of the \( CLD \) haplotype from her mother, including \( MET \) and all markers distal of it. However, the proximal part of her maternal chromosome including \( D7S501 \) and \( D7S496 \) derived from the non-\( CLD \) haplotype of her mother; \( D7S523 \) alleles could not be unambiguously assigned. Therefore, the normal allele for the \( CLD \) gene segregated with the proximal markers \( D7S501, D7S496 \), and, possibly, \( D7S523 \). \( CFTR \) was excluded as a candidate for \( CLD \) by this recombination.

**DISCUSSION**

Linkage disequilibrium as a mapping tool for genetic diseases may have considerable power in selected populations. It may be possible even in rare diseases to imply or exclude candidate genes by studying polymorphic markers that are sufficiently close to them. This is the case for \( CLD \) in the Finnish population where the disease shows elevated incidence and geographical clustering suggestive of a founder effect (8, 22, 23). In \( CLD \), the rarity of the disorder made it difficult to obtain enough families for conventional linkage studies with RFLP markers, but the exclusion of candidate genes seemed possible by considering allele or haplotype associations as well as linkage in a small number of families. We chose to start with three genes that were more or less suggested as candidates by their involvement in chloride transport. Each of the candidates turned out not to be the \( CLD \) gene but the \( CLD \) gene was mapped close to one of them—whether this result was just serendipitous or perhaps reflected the clustering of related genes remains open.

The study was initiated with \( CFTR \), for which good polymorphic markers were readily available. When preliminary analyses indicated no haplotype or allelic associations but also failed to show recombinations, the studies were extended to the region around \( CFTR \) in \( 7q \). Highly informative microsatellite markers that had become available confirmed genetic linkage, and significant allelic associations were detected with two markers located proximal of \( CFTR \) in \( 7q31 \). Haplotype analyses formally excluded \( CFTR \) as the \( CLD \) gene.

The population history of Finland (22–24) and the almost exclusive clustering of \( CLD \) in a region of Finland where very few inhabitants lived before the 17th century and where the population then grew starting from a relatively small number of founders (8, 22, 23) suggest that the gene expanded <400 years ago. The observed strong allelic association suggests that most \( CLD \)-carrying chromosomes were derived from one founding mutation. The relatively short time reduces the
number of meioses separating the founder and present-day patients, causing linkage disequilibrium to be expected at longer genetic distances than in, e.g., diastrophic dysplasia or cartilage-hair hypoplasia (25, 26). On the other hand, 400 years have preserved the allelic association even for microsatellite markers that may have relatively high mutation rates (25). Although the present data are numerically limited and, therefore, insufficient for the estimation of the genetic distance between D7S496 and CLD, the frequency of allelic I and 5 in normal and CLD chromosomes is compatible with the hypothesis that the original CLD mutation occurred or was introduced on a chromosome carrying allele 6 for D7S496 and that only one historical recombination occurred with the most common allele 1. Similarly, one can hypothesize that the D7S501/D7S496 haplotype was 6/6 in the chromosome(s) in which the mutation was introduced; haplotype analyses did not allow the construction of a more extended ancestral haplotype. More accurate estimates may become possible by studying allelic associations in more patients and by applying the Luria–Delbrück analysis (as in ref. 25) with extended data. This approach is more powerful than linkage analysis because families with a single affected child are fully informative.

Both known human Cl−/HCO3− exchange protein genes, AE1 (at 17q12–q21) and AE2 (at 7q35–q36), seem to be excluded as candidates for CLD by these results. The formal exclusion of AE2 as the CLD gene must await the results of further linkage studies, even though its cytogenetic localization four chromosome bands away argues against its candidacy. There are examples of structurally or functionally related genes being clustered to regions of hundreds of kilobases or to the same chromosome (e.g., homeobox genes, blood group antigen genes) as well as spread over the genome (e.g., collagen genes). The CLD gene maps close to CFTF, AE2, CLCN1 (a skeletal muscle chloride channel gene at 7q35, ref. 27), and ABPI (an amiloride-binding protein gene involved in the regulation of sodium trafficking in kidney, at 7q33–qter, ref. 28). This result may suggest the clustering of genes involved in ion trafficking in the long arm of chromosome 7. Interestingly, a mucin gene (MUC3, ref. 29), expressed primarily in the colon, is genetically linked to CFTF. Mucins are highly glycosylated protein components of the apical membranes of secretory epithelial cells, but neither their structure nor known functions suggest that they might be involved in CLD. Rather, the association may represent the clustering of tissue-specific genes.

At least eight Cl−/HCO3− exchange protein genes and highly homologous genes from different species are known, encoding exchange proteins in human, mouse, rat, chicken, and trout. The alignment of these sequences revealed highly conserved regions, suggesting possible functional significance. Thus, a strategy to isolate candidates for the CLD gene can be formulated. The region around D7S496 can be isolated in yeast artificial chromosomes; in current yeast artificial chromosome libraries, a screening typically yields several clones covering 400–800 kb, and one round of walking to both directions from the original clones can yield contigs of 1.5–2 Mb (30). The human inserts can be analyzed by PCR and hybridization using oligonucleotide primers and probes corresponding to highly conserved segments in anion-exchanger genes (cf. ref. 31). If no candidate genes are found by this approach, other strategies capable of detecting many types of genes from cloned genomic DNA, such as the direct selection of cDNAs (32, 33), can be used. The CLD gene may or may not be a structural gene for a transport protein, but its role remains speculative until it is molecularly cloned and characterized.

We thank Ms. Pia Höglund for excellent assistance in the laboratory and Dr. Eric D. Green for the gift of primer solutions for many PCRamarkers. This work was supported by the Academy of Finland and the Sigrid Juselius Foundation. Part of the study was done at the Folkhälso Institute of Genetics.